

DOCUMENT RESUME

ED 034 295

EA 002 606

AUTHOR Fearing, Joseph L., Ed.; Kowitz, Gerald T., Ed.
TITLE Some Views on Longitudinal Inquiry. Research and Services Series.
INSTITUTION Houston Univ., Tex. Bureau of Educational Research and Service.
REPORT NO RSS-321
PUB DATE Jul 67
NOTE 194p.
EDRS PRICE EDRS Price MF-\$0.75 HC-\$9.80
DESCRIPTORS Developmental Programs, *Educational Research, Evaluation, Evaluation Techniques, Learning Processes, *Longitudinal Studies, Research Criteria, *Research Methodology, *Research Problems, Statistical Analysis, Taxonomy, *Time Perspective

ABSTRACT

This document provides a current assessment of the role and character of longitudinal studies. Eighteen articles define this type of study, discuss its applications to the educational process, advise when to use longitudinal inquiry, and describe barriers formed by lack of statistical techniques in manipulating the data. The traditional argument involving the major variable of time in longitudinal studies, with the uncontrollable dimensions it imposes on the study group, is examined and some novel ideas are proposed. One such idea is the unanticipated changes that occur to the researcher himself as he acquires new methodology. Another is an application of statistical assumptions to the study group by approaching the subjects as random variates, thereby overcoming the loss of data from loss of subjects in the group. The lack of knowledge about classroom anatomy adds unidentified variables to longitudinal studies even under ideal conditions. Therefore, research on this subject must progress in order to strengthen longitudinal studies. In general, the respondents feel that this type of study provides a broad framework within which all of the tools of science can be used. (LN)

ED034295

SOME VIEWS ON LONGITUDINAL INQUIRY

Editors:

Joseph L. Fearing

Gerald T. Kowitz

A Publication of the
BUREAU OF EDUCATION RESEARCH AND SERVICES

University of Houston
Houston, Texas 77004

1967
U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE
PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION
POSITION OR POLICY.

RESPONDENTS ON LONGITUDINAL INQUIRY

Stuart A. Courtis, Michigan
Warren Coxe, New York
Robert L. Ebel, Michigan
Jack R. Gibb, California
Harry F. Harlow, Wisconsin
Glen Heathers, New York
Joseph A. Johnston, Michigan
Phillip Kraus, New York
Robert Lathrop, Pennsylvania
Eric Lunzer, England
Frederic L. Matthews, New York
Harold Mitzel, Pennsylvania
Ralph Ojemann, Ohio
Hanus Papousek, Czechoslovakia
A. Harry Passow, New York
Robert A. Peck, Texas
William Rabinowitz, Pennsylvania
Hans Selye, Canada
Ruth Strang, Arizona
J. Wayne Wrightstone, New York

PREFACE*

Since the school, and thereby the educative process, accounts for most of the time from late childhood through young adulthood, the processes of changes over time are salient. Because of the long periods of time involved, studies which do not account for the time factor are often criticized. Since education is irrefutably a longitudinal process, the appropriate procedures for studying it should be longitudinal. Latitudinal studies which present a picture of only an instant of time, must by definition fail to account for some of the most critical aspects of the educative process. The compromise of creating a longitudinal study by placing a series of latitudinal studies in sequence is criticized by both the proponents of longitudinal studies as inadequate and by the proponents of latitudinal studies as unnecessary.

Some Traditional Arguments

In spite of its apparent appropriateness, the longitudinal approach has been more of a bone of contention than a rallying point. Furthermore, while the idea of longitudinal studies seems to be an excellent one, few good examples of such studies exist. The most common criticism is sampling. To the extent that the groups are equivalent and the data receives equivalent treatment, the results should be equivalent. However, the problem of equating groups remains today as a major problem in research theory. When viewed as a problem of sampling, longitudinal studies are criticized as being nothing more than latitudinal studies repeated periodically. Even seriatim studies can be so described.

A practical problem arises from methods of financing them. Most agencies which grant funds for research are interested in results within a defined period; they do not like to commit resources into the distant and nebulous future.

* Adapted in part from a paper presented to Session 94: Research Design in Relation to Theory of Growth Relationships, 1966 Annual Meeting of the American Education Research Association, Chicago, Illinois.

Another related criticism is that at the end of a longitudinal study, the conditions have changed sufficiently that its beginnings are no longer valid. When the continuity is lost, so is the point of doing a longitudinal study.

Another major difficulty, not unrelated, is the problem of sustaining a sample. While the children in school appear to be available to the research group, neither the school administration nor the researchers can control the mobility of the child's family. Thus, the sample may be depleted or at least severely damaged long before the span of time planned for the study has elapsed.

The passage of time also has an impact upon the techniques used in a study. As knowledge develops, new techniques emerge, and new theories develop which present alternatives to the original plan. Archaic instruments cannot be justified merely because they were the ones with which a study was started; atavism is a poor stimulus for progress. Changing instruments during a study introduces unnecessary confusion since the introduction of new alternate measures cannot but destroy some of the longitudinal aspects of the study. The Quality Measurement Project of New York State faced this problem because of changes in the Iowa achievement tests which it used as a basic instrument.

Perhaps the greatest difficulty in longitudinal studies is in the processing of data. In longitudinal studies time is presumed to be a major variable. Few statistical techniques can accommodate this assumption. In fact, many of them specifically rule out time as a significant variable. Chi-square, for example, cannot be used legitimately to study pre-post test data and, of course, the pre-post test design is an elemental longitudinal study.

In the face of these commonly recognized problems, it seems necessary to have a careful assessment of the nature and role of longitudinal studies. It may be significant that the criticisms refer to "practical" problems and most superficial ones. There seems to be no fundamental theory against which a meaningful attack can be mounted or from which the idea can be defended. This is, perhaps, the most serious of all criticisms.

The Past

In the past, longitudinal studies have provided mostly descriptions of observed phenomena. Classic examples are the Gesell development schedules, and curves of learning and physical growth as presented by Curtis, Olson, Wetzel and others. These are all attempts to describe patterns across a span of time. Some workers verbalized a hope that when workable descriptions were found, meaningful classifications could then be developed which would allow carefully controlled experiments of growth and development. The hopes are rudimentary in comparison to the goals currently espoused by multivariate (and essentially latitudinal) research designs which intend to achieve prediction and control.

In spite of the history of longitudinal studies, little theory has developed either for their conduct or for the use of the results. There is little that exceeds the primitive, action oriented patterns on the managerial level of the executive hierarchy; the idea seems to lack heuristic power. But in spite of all criticism, the spirit refuses exorcism.

A Look to the Future

Strangely enough, perhaps the great impetus for the recent development of longitudinal techniques did not arise from within the human or behavioral sciences. Instead, it has grown through the needs of big business to maintain inventory control over extended periods of time. Also, industrial production has had to become a long-term affair and has had to develop models for time-linked planning (e.g. PERT, CPM, etc.). The planning for the manufacture of atomic submarines and space vehicles has brought attention to the problem of evaluating development through time. (Admiral Rickover may one day be hailed as the stepfather of longitudinal studies in education!) Another recent advance, the electronic computer, has provided the ability to process data through mathematical models, which in the past, could not be approached because of their complexity.

Since one basic variable in a longitudinal study is the passage of time, and since this variable is so obviously

critical in education, it seems reasonable to suggest that new attention will be given to the approach. New techniques will be developed which will overcome or at least mollify many of the problems which have traditionally plagued the approach. It seems not unreasonable to predict that with the amount of funds currently being made available for research in education, that some time in the near future a major effort will be made to develop theory and mathematical models suitable for longitudinal research and subsequently to study the longitudinal problems which are inherent in the educative process.

The Study

As a means of further examining some of the advantages and the problems of longitudinal methodology, eighty-five recording tapes were mailed to researchers who were believed to hold either a hostile or a friendly attitude toward the notion of longitudinal studies. Not all were from education or related disciplines; six reside in foreign countries.

In an attempt to gain some uniformity in the comments, the following suggested outline for their remarks was included:

State your Name, Institution and Date

Section One:

Please reflect on your own research and experiences, giving opinions on the role of longitudinal methods in your research. You need not organize or formalize these reflections; your informal comments are sought and appreciated.

Section Two:

From your viewpoint:

1. What are the salient characteristics of longitudinal research?
2. What important theoretical assumptions are unique or indispensable to the longitudinal approach?

3. What role have these assumptions (item 2) played in your use of the longitudinal approach?
4. What statistical or mathematical models are most helpful in employing the longitudinal approach?
5. What is unique to the longitudinal approach which limits the use of certain common statistical techniques?
6. What contributions may longitudinal studies make to each of these scientific hierarchical levels?
 - a. Classify--describe
 - b. Understand--explain
 - c. Generalize--theorize
 - d. Predict--control
7. What major barriers have you faced in using the longitudinal approach?
8. What is the general promise of longitudinal research?

Not all of the respondents chose to follow the outline. Some invited colleagues to join them in the recorded discussion; others preferred to write rather than to record their comments.

Twenty-seven responses were returned. Systematic content analysis did not seem to yield the valuable information contained therein. Requests, therefore, were made to publish twenty-one of these responses in full. Twenty agreed to do so.

With very few changes these transcripts provide the substance of this monograph. This publication was produced in the belief that such spontaneous reflections by these eminent persons afford a dimension of considerable value to the student of research methodology not commonly found in standard references.

Discussion

Certain points emerge. If one were to group the responses in terms of the value they placed upon longitudinal methodology, one would arrive at categories which could be labeled "yes--maybe," "No--maybe" and "Maybe--maybe." Few were willing to take a position without either antecedent or subsequent qualification. All recognized that there are

significant methodological problems. Several pointed out that some investigations must be longitudinal because of the nature of the problem; other problems can be approached only through a latitudinal design.

It is generally recognized that longitudinal methodology has not developed the sophistication in technique that marks the statistics of latitudinal studies. Some persons view the development of more exotic techniques as an urgent need while others observed that the less restrictive framework of the longitudinal study was one of its great advantages. It permits the researcher to adjust his strategies as new evidence develops. The feedback from a cybernetic system should lead to improvement of search methods.

The use of the same subjects throughout an investigation with an extended time axis emerges as the sine qua non of longitudinal inquiry. One respondent pointed out that if proper statistical generalizations are to emerge from the findings, the subjects as a group must be approached as another of the random variates.

Several of the respondents were sufficiently concerned about the labeling of different methodologies that they attempted to develop taxonomies of research strategies. Some that were suggested are Prospective--Retrospective; Natural Experimental Variatiabiles--Induced Experimental Variables. Perhaps the most interesting contrasted the Longitudinal with the Shortitudinal, rather than with the traditional "cross-sectional" or the isomorphic "latitudinal" labels.

Most of the difficulties which were reviewed in the first part of the paper were recognized by the respondents. Several additional problems seem worthy of mention.

In a longitudinal study, we must anticipate changes in the researcher himself. Certainly a competent researcher will be acquiring new methodology constantly and presumably will apply the best technique available to the problem at hand. In the same vein, it is probably inevitable that he will experience changes in attitudes which may modify some of his activities in the study.

A problem that may be unique to longitudinal research in the schools is that the anatomy of the classroom is largely unknown. As a result, experimentally induced variables may be negated or at least weakened by more potent but unidentified variables which exist unrecognized in the classroom.

These, then are some of our first impressions of the data. The study will continue. In general, it seems that the respondents perceive a longitudinal approach as a grand strategy rather than as a specific tactic. As such, to speak of longitudinal methodology may be inappropriate. As a strategy, it would not be limited to specific techniques but rather it would provide a broad framework within which all of the tools of science would find a place.

Without attempting to summarize further we submit these responses to you the reader for your consideration.

We want to express our thanks to the many important people who took time to respond to our request, especially these persons who so graciously granted permission to publish their responses, to the University of Houston for the grant which made this study possible, and to Mrs. Marjorie T. Turko for typing and production related assistance.

Joseph L. Fearing
Acting Director
Bureau of Education Research
and Services
University of Houston
Houston, Texas

Gerald T. Kowitz
Director of Research
Mid-America Center for Urban
Education
University of Oklahoma
Norman, Oklahoma

July, 1967

TABLE OF CONTENTS

TITLE	PAGE
TOWARD A DEFINITION OF LONGITUDINAL METHODOLOGY	
by Stuart A. Curtis	1
LONGITUDINAL INQUIRY AND THE INDIVIDUAL	
by Warren Coxé	5
LONGITUDINAL RESEARCH--A DISSENTING OPINION	
by Robert L. Ebel	9
COMMENTS ON LONGITUDINAL METHODOLOGY	
by Jack R. Gibb	28
SOME SALIENT CHARACTERISTICS OF LONGITUDINAL	
RESEARCH by Harry F. Harlow	33
THE STUDY OF SCHOOL PROGRAM CHANGES OVER TIME	
by Glen Heathers	36
LONGITUDINAL INQUIRY: SOME REFLECTIONS	
by Joseph A. Johnston	45
MY REFLECTIONS IN RESPECT TO THE APPLICATION OF	
LONGITUDINAL METHODS TO THE EDUCATIONAL PROCESS	
by Eric Lunzer	48
SUBJECTS AS A RANDOM VARIATE IN LONGITUDINAL INQUIRY	
by Frederic L. Matthews	61

TABLE OF CONTENTS

TITLE	PAGE
THE LONGITUDINAL APPROACH TO RESEARCH WITH CHILDREN:	
A DISCUSSION by Harold Mitzel, William Rabinowitz, and Robert Lathrop	72
SOME REFLECTIONS ON MY OWN RESEARCH AND EXPERIENCES	
WITH LONGITUDINAL METHODOLOGY by Ralph Ojemann . . .	101
WHEN LONGITUDINAL INQUIRY IS ESSENTIAL	
by Hanuš Papoušek	108
LONGITUDINAL INQUIRY IN RETROSPECT	
by A. Harry Passow	121
COMMENTS ABOUT LONGITUDINAL RESEARCH	
by Robert F. Peck	137
ON LONGITUDINAL INQUIRY by Hans Selye	146
SOME IDEAS ABOUT LONGITUDINAL INQUIRY	
by Ruth Strang	149
A DIALOGUE ON LONGITUDINAL RESEARCH METHODOLOGY	
by J. Wayne Wrightstone and Phillip Kraus	153
A BIBLIOGRAPHY OF MEASUREMENT OF GROWTH; FUNCTION	
FLUCTUATION by Frederic M. Lord	181

TOWARD A DEFINITION OF LONGITUDINAL METHODOLOGY

Stuart A. Courtis
Professor Emeritus of Education
University of Michigan

A term usually gets into circulation through its use by some agent with a special meaning. Thus "longitudinal" has no established connotation. My first "longitudinal" study was made about 1900, although if anyone called it that I would not have known what they meant.

To me the salient characteristics of longitudinal research are, first of all, repeated measurement of the same subjects year after year. However, I might use the term even if different subjects were used each year, if the study was instituted to appraise the change in scores over successive periods of time. Today I would not like to use the term unless identical subjects were measured at each appraisal over successive periods of time and the scores of each individual subject were preserved, and not combined statistically with those other individuals.

The theoretical assumptions that are indispensable to the longitudinal approach are: (a) belief that individuals are unique, no two alike, (b) the purpose to discover the changes in each individual over successive periods of time, (c) any massing of scores hides the individual variations one

wishes to study, (d) that raw scores may be "corrected" for factors previously proved to be operating, (e) that trends if research has established the laws of operation of this factor are expressed as the number of individuals alike in every known factor to determine whether or not there are unknown factors also operating.

What has been the effect of my use of the longitudinal approach upon my understanding of the purpose of longitudinal research? It has constantly improved my understanding of the purposes of research and has clarified inferences drawn from the results.

Any score, or combination of scores, is a statistic. If one uses the best model he knows, and then is critical of his scores, the longitudinal method constantly exposes the inaccuracy or incompleteness of his results. The longitudinal approach as a basis of criticism is the most powerful influence for progress I know.

The unique feature of longitudinal approach is that it exposes the inadequacy of results obtained by any other statistical method. Mass statistical results without exception, (from my point of view), give rise to conclusions which are obviously inadequate and therefore false.

In my thinking the order of classify--describe, understand--explain, generalize--theorize, predict--control would be exactly reversed. Most important is prediction and control, generalize and theorize, understand and explain.

Regarding the hierarchical levels to which longitudinal inquiry may make contributions, let me add another question: What does longitudinal research contribute to the definition of reality?

Most persons I have questioned, after thinking about it, agree that any three dimensional object is real. If I ask them, "Is love real?" they admit it is, but become so confused that they ask for more time to think it over.

If we examine the course of evolution, we tend to analyze its progress into stages or levels: energy, matter, life, conscious, mind, organization of mental ideas and peoples, and intangibles like love, beauty, etc. However, analysis from a different viewpoint reveals that at each stage, evolving organism like a plant or animal exhibits some behaviors which are characteristic of the next higher stage. Such types of behavior cannot be defined in terms of the characteristics of a given level.

For instance, a plant is not supposed to have intelligence. But a root, meeting a rock, twists around it and goes on down deeper after water and food. This is intelligent

behavior. A dog associating with humans apparently comes to feel affection for them. A human being, under certain conditions not under his control, becomes clairvoyant. There very clearly is another level of life beyond those we recognize.

The supreme function of longitudinal research is to direct our attention to that higher level and reveal its characteristics.

LONGITUDINAL INQUIRY AND THE INDIVIDUAL

Warren Coxe
Director of Educational Research (Retired)
New York State Education Department

I am interested in longitudinal studies because of the light they may throw upon individual differences. Individual differences in education are important, partly because of their value to guidance, partly as an aid to forming groups in school for purposes of instruction and partly to aid in the construction of a curriculum suitable to pupils of different needs.

Longitudinal data may be of various kinds. We may use the measures of ability as obtained from general intelligence tests or from measure of achievement or from measures of special abilities. The first is illustrated by such tests as devised by Binet and Terman, the second by such tests as the Stanford Achievement Tests, and the third by the Primary Abilities Tests devised by Thurston.

The Division of Educational Research in the New York State Education Department has used the data from the Harvard Growth Study. This represents measures of general ability. What growth patterns would be manifest were other kinds of tests used we have no suggestion. One may guess

that were we to use Spearman's concept of general intelligence we would look for growth of the general factor and also of specific factors. Were we to use Thurston's Primary Abilities Tests we would supposedly get different curves for different abilities. What work we have done makes me very curious to try different measures than we used to see whether the growth curves are similar or different.

The growth curve may well be considered a method of presenting the final result of the use of certain measurements. Its use as I see it now is confined to classification of these curves and where valid differences are found to give their interpretations.

It seems possible, however, that these curves might throw light upon the relative importance of inherited abilities and environmental conditions.

One difficulty in all longitudinal studies is that the data must be collected over a period of probably ten to twelve years. This makes the study very expensive. The Harvard Growth Study data were used by us just because they were available. We did not gather independent data to make the study. The individual growth curves were drawn from these data. They can be studied for items of similarity

and dis-similarity. If they can be classified into groups some interpretations can then be suggested for school use.

We were faced with another problem; namely, that of establishing units of growth which could be used in plotting the curves. In other words, we needed to use a unit of growth for the younger ages which would be the same as a unit of growth for the older ages. The usual handling of test data is to get data on subsequent years or subsequent grades, then note the differences. In this case we were not doing that but rather we were using the same pupils and taking records of them year after year. In other words, we were faced with the problem of making the units of the scale equal at all ages. In order to do this, Dr. Cornell and Mr. Armstrong worked out a unit which they called CMC, CMA. This is not repeated here because it appears in their original publication.

One may wish to get a single measure for growth. This does not seem possible to us when we are dealing with longitudinal studies. We can, however, get a curve, and curves become the unit of consideration. This may be considered when we attempt to do any predicting. We do not have a satisfactory method but it seems possible that the first few measures on the curve will be roughly predictive of the shape of the curve.

The major barrier which is faced in using the longitudinal approach is gathering the data. It's a very difficult and time consuming job to get data on the same pupils over a period of ten or twelve years. Now, I say ten or twelve years, but it is possible that time would be significant if it was less than that, but we do not have the evidence. What we have used is the Harvard Growth Study which runs to twelve years. And I think it is necessary to have a number of years in the elementary school, and a number of years in the high school in order to get any great value from such a study as the longitudinal ones. However, the general promise of longitudinal research seems to be very great. I think when we really develop it, we will have developed an instrument which is very much the superior of any instrument we have now in understanding pupils, and therefore, getting suggestions for improved school programs, school curricula, and school organization.

LONGITUDINAL RESEARCH--A DISSENTING OPINION¹

Robert L. Ebel
Professor of Education
Michigan State University at East Lansing

I was involved in some of the planning of the growth study which is currently being conducted by the Educational Testing Service. As you may know, the purposes of that study were to develop some growth norms for the Sequential Tests of Educational Progress and to study some of the factors that influence educational growth. I found myself in the role of critic quite frequently during the planning of this study.

First of all, the concept of "growth" as applied to educational development seems to me not entirely appropriate to educational development. The processes of biological growth in plants and animals are largely automatic and self-regulated. It is true that man can change them by altering conditions, but ordinarily they take place and go on once life has been started, if the proper conditions exist for it.

¹My direct participation in longitudinal research studies has been limited. All of the studies that I have been concerned with, directly, have been cross-sectional, one-shot studies.

In contrast, it seems to me, the processes of formal education are not automatic. Their success, the speed at which they progress is highly responsive to changes in conditions. As a matter of fact, all teachers, it seems to me, are constantly at work, trying to discover or invent conditions which will facilitate educational progress. So growth in learning is somewhat different from growth in the biological sense.

Scarvia Anderson, in reporting on this growth study in the Journal of Teacher Education a few years ago, pointed out that occasionally there were declines in test performance so that the term "growth" didn't seem entirely appropriate for cases of that sort; but she rejected the term "change" because she said, "Actually, it's growth the schools are interested in producing." In that sense, I have no quarrel with use of the term "growth." What I do question is the effort to obtain growth norms as if intellectual growth, mental growth were something which goes at its own rate, whose progress we can discover and record, and which, therefore, is something to be expected regardless of the methods used in instruction.

This seems to me to be stretching the concept of

growth a little too far. For the process of education is one that is highly subject to human control, the control of the teacher, the control of the learner, and the control of the community that provides the educational resources. Professor Robert Green, my colleague at Michigan State University, studied the effects of educational deprivation on the children of Prince Edward County in Virginia. He has extensive data to support the conclusion that the process of education is not one that goes on at its own rate regardless of educational efforts. It is highly sensitive to the conditions surrounding it, to the stimulation that is provided. In that kind of situation it seems to me a little inappropriate to talk about educational growth. I don't think this is just a trivial quibble over terms. It seems to me to call into question any kind of a long-term longitudinal study that aims to find out how children "grow" educationally. The concept of growth as it is applied to longitudinal studies is one that I think may lead our thinking astray, and may cause us to tend to underestimate the importance of paying careful attention to the conditions under which growth occurs.

The method studies which aim to find out how this particular resource of that particular method affect learning

are a different kind of study. They relate methods to results. They are not what we ordinarily include under the heading of longitudinal studies, although, of course, they do involve measurement at more than one point in time.

The argument for longitudinal studies as opposed to cross-sectional studies is that you can't get a good picture of how a 9th grader in 1965 is going to look as a 12th grader in 1969 by studying 12th graders in 1965. But the problem with a cross-sectional study is that these children at different stages in their development go through different kinds of experiences. We may sometimes forget that the very thing which spoils our cross-sectional study as a basis for longitudinal interpretations also would spoil the results of a longitudinal study as a basis for inferences about the behavior of children far into the future. Each one of us, each age and generation, lives his life in a different world. Our circumstances are different. The world changes.

All of our elaborate experimental designs, sampling studies, etc., have not given us any way of sampling the future, or of drawing valid experimental inferences with respect to it.

We don't even sample the past very often. But there is absolutely no way in which we can sample the future. It sometimes seems a little strange that we are so willing to make inferences about the future performance of students on the basis of so little valid data. We worry greatly about representativeness in other aspects of sampling. There is one that is really crucial, that we don't sample at all simply because we can't get at it.

Coming back to the growth study at ETS and its interest in growth norms, I have sometimes wondered what form these norms would conceivably take. Perhaps if a student in the fifth grade gets a particular score on the Step Test, growth norms might report his chances of getting another, higher score in the seventh grade. You might have a two dimensional expectancy table of scores at 5th grade against scores at 7th grade. But as I think of the elaborateness of these tables for each of the tests, I then begin to wonder who would be really interested in all these norms. I wonder, too, how accurate and relevant they might be in particular cases.

How important is it to be able to say on the basis of a student's score on one test, at one level of his educational development, what he is likely to get when he is tested

with a similar test two or four or five years later? Remember, too, that these norms will inevitably be based on a particular system and pattern of education that was sampled when the norms were devised, and that this may be quite different from the situation that obtains when the individual is tested in the future. These kinds of questions have led me to wonder a little about the actual utility of what must necessarily be a fairly elaborate set of growth norms.

Returning now to the general notion of longitudinal studies, let me report that I have encountered on several occasions, special exhortations from research workers who have done some longitudinal studies or who have been critical of cross-sectional studies. Their exhortations are to the effect that we ought to be doing more longitudinal studies. They argue that we will never really get anywhere in understanding human development until we follow individuals throughout their careers. But never have I encountered any very clear rationale for this point of view; no very clear defense of this as the necessary way to make progress in the understanding of human behavior. As a matter of fact, there are some reasons to believe that this method has some serious built-in drawbacks. I do not propose to dwell on the familiar ones at this point:

the difficulty, the cost, and the time requirements that we are all familiar with. There is another serious problem, it seems to me. Whenever you select a particular sample of students and begin to test them extensively, or follow their lives carefully, accumulating special kinds of records on them, you can't very well do this without affecting them. There may be a Hawthorne Effect when they are singled out for special kinds of study. Just as the physicists are never able to demonstrate what their theories assume with respect to the behavior of an isolated particle that is free from the influence of any other nearby particles (because in order to measure its velocity or its position, you must always bring some measuring instrument into the field) so even to a greater extent, in educational research, in order to do an intensive longitudinal study, you have to affect the lives of the individuals being studied. Thus they are no longer typical; they are not the same as the unstudied individuals. Yet we would like to make inferences from our findings on them about other individuals who haven't been studied. This, it seems to me, poses some fairly serious problems.

A couple of British psychologists have studied the influences upon students of repeated testings with the same

tests or the same kind of test. Their finding was that the students kept on improving their performance in the test even up to ten repetitions of the test.

Testing instruments (almost always we do use some sort of a testing instrument in making longitudinal studies) also tend to wear out with use; they tend to get out of date. These famous then-and-now studies in education, where tests that were developed in one educational period are readministered in a much later period, almost always turned up items which are no longer appropriate. They are no longer relevant because conditions have changed. This is another difficulty in the way of longitudinal studies.

In the face of all these difficulties, it seems a little strange that apparently nobody has settled down to do a thorough-going rational job of defending the importance of longitudinal studies. The virtues of such studies seem to be taken for granted. When a particular study doesn't work out well, the reason often given is that the investigator wasn't able to study his subjects over a long enough period of time.

One reason human beings are as effective as they are in the world is that they adapt. Their behavior tends to change with changing circumstances. It is different for

different people. But this very adaptability makes them rather poor subjects for rigorous scientific study. The thing that you are studying about them doesn't stay the same. Part of the reason why it doesn't stay the same is that you are studying it. All of this, I suppose, leads to an observation that might be challenged. But it is my observation that the most useful longitudinal studies that have been reported are those dealing with physical growth and development of individuals, not with their mental development.

There have been some fairly extensive and important attempts to study individuals longitudinally. But at the moment, I can't think of any of them that have really added important new insights; that have made any permanent difference in the way in which we are handling the educational process. It can hardly be that people have been neglecting the important findings, or have simply refused to apply them. It seems to me that perhaps the findings have not been very important. I come back again to the notion that we need a strong rational defense for the importance of longitudinal studies. I can see a number of rational arguments against them, some of which I have mentioned or hinted at here. If we are going to promote them as an essential part of educational research some strong defense of them would be highly desirable.

In this connection as I try to buttress my own casual information about longitudinal studies, I find that few of them seem to be reported. Not many studies in research methodology have to do with longitudinal research techniques.

The Encyclopedia of Educational Research has references in just a few paragraphs to longitudinal research studies. In the article on research methodology in the Review of Educational Research, I found nothing that would contribute to this presentation at all. Articles on research methodology dealing with longitudinal studies in the Education Index are few and far between. This might be interpreted as an indication that research workers have been neglecting a very valuable and powerful tool. On the other hand, it might indicate that, when you get right down to cases, the tool is not nearly as powerful and fruitful as its proponents claim. In terms of time, cost, and results yielded, it seems to be a particularly attractive research tool.

I wouldn't want to carry this line of argument so far as to seem to suggest that we can't learn anything from history. Actually, a longitudinal study is essentially a historical study. I think we can and we do learn things from history. But the variables are so numerous and so difficult to control that scientific research models which call for

control or randomization of all of the relevant variables simply don't work very well. We are driven back into what Campbell, Stanley and others who have referred to as quasi-experiments in which our interpretations must be guarded because we haven't been able to satisfy many of the critical assumptions. In these studies the dividing line between scientific research on the one hand and practical experience on the other becomes rather fuzzy.

I suspect that most of what we do that is good in education, is the consequence of practical experience rather than of scientific research, i.e., of rigorously controlled experimental studies, I don't want to imply that there is no place for rigorous scientific research in education. But my present belief is that we need to place as much or more emphasis on records of practical experience. These are unique in one sense, yet they are to some extent repeatable by others. We need to place as much emphasis on the accumulation of knowledge about how to educate as we do on experimental, well-controlled research studies.

This is about all I ought to say on this subject. It may be a good deal more than I should have said, in view of the limitations of my own direct involvement in longitudinal

studies. It relates to what seems to me to be basic issues about the role of research in education, and about the kinds of research that are likely to be productive. It carries some philosophical implications about the suitability of human behavior as an object of rigorous scientific study. But this probably is not the time or place for a long dissertation on that topic.

What are the salient characteristics of longitudinal research? It involves repeated measurement or observation of the same individual, over a period of time. How long that time should be may vary widely. The typical longitudinal study involves the same individuals or groups over a number of years.

What important theoretical assumptions are unique or indispensable to the longitudinal approach?

I'm not sure that the longitudinal approach per se involves any special theoretical assumptions. It is true that when we begin to apply particular methods of data analysis, we often need to make special theoretical assumptions. Multivariate analysis and factor analysis of three-dimensional data matrices have been developed partly in response to the characteristics of data that are obtained in longitudinal studies.

One obvious characteristic of longitudinal data is that the measures for the same individual at different points in time are likely to be correlated. Individuals who score high initially are likely to continue to score high and those who score low are likely to continue to score low. There is no reason to assume that measurements of the same person are independent. This co-variance of observations at different points in time places some restrictions on the methods of statistical analysis that are used. This is one necessary theoretical consideration.

Then further, the justification for longitudinal studies rests at least partly on the belief that if we can study adequately a sample of individuals over one period of years, the findings will be useful in some future period of years. This assumption was mentioned previously. It is one which may not be completely valid. Whether it is invalid enough to spoil the results of the experiment may depend on what is being studied, how labile it is, how much it is affected by immediate circumstances. One doesn't have to be very old to realize that educational climates change. One current educational problem for example, is how to get a youngster into college. Quite apart from the financial problem there is the problem of just getting him admitted into the college of

his choice. It wasn't nearly so much of a problem 20 years ago. It may not be nearly so much of a problem in the future. These changing conditions do tend to limit the validity of inferences that are made on the basis of studies at one point in time. Even a longitudinal study is, in a sense, always time-bound. It is always determined in part, by the conditions that exist at the time when the study was made.

What role have these assumptions played in my use of the longitudinal approach? I haven't used it, so for me the assumptions haven't played any role. The difficulties inherent in longitudinal research are the main reason I haven't used it. There are other reasons, among which is the fact that the problems I've been interested in didn't seem to require it.

What statistical or mathematical models are most helpful in employing the longitudinal approach? I don't feel competent to answer this question. A symposium at the University of Wisconsin several years ago, whose proceedings were edited by Chester Harris, was entitled "Problems and the Measurement in Change." In it a number of theoretical techniques and mathematical models that would be useful in longitudinal studies were described. I must confess to some feeling that the mathematical tools have been elaborated far more than the logical foundations of the method, and the

quality of the data that we extract from it would warrant. This is not to criticize statistical elaborations. It is not their fault that things change as rapidly as they do, or that the data are no better than they are. But perhaps it is to say that inadequacy or limitations of statistical techniques or mathematical models may not be a major obstacle to productive longitudinal research. The difficulties seem to me, to lie mainly in other areas and other directions.

What is unique to the longitudinal approach which limits the use of certain common statistical techniques? The lack of independence of the observations has already been mentioned to be taken account of in statistical tests of significance. There are probably other unique features of the longitudinal approach which limit the use of common statistical techniques. Simple difference scores between pre- and post-tests, or early and later measurements, are likely to be highly misleading as measures of change because of their neglect of regression effects.

In terms of the following scientific hierarchy, where can longitudinal studies make the greatest contribution? These four categories are: classifications or description, understanding or explanation, generalizing or theorizing, and prediction or control. I cannot answer this question in a way

which singles out longitudinal studies from other research studies. It seems to me that research is likely to contribute most to education through the description of things as they now exist. Certainly those descriptions can be generalized, and lead to an understanding at the level of what goes with what, and what precedes what. Research can contribute most in those areas.

I'm quite skeptical, in view of the complexity of the process of learning and the plasticity of human behavior, that we're likely to get any very powerful theories or as far as individuals are concerned, I doubt that we will ever make very accurate predictions, or exert very much control. There may be an ethical issue here too--whether human beings out to be subject to control by other human beings once they've attained maturity--on anything other than social legislation or social custom. In other words, I doubt that a research worker ought to find out something about people which will enable him to control their behavior. So my answer to this question is that longitudinal studies as well as other research studies are likely to be most productive when they fill gaps in our knowledge about the way things are.

How many students are dropping out of high school

during high school years? How much demand is there for mathematics teachers? These are, I realize, very low level questions that many research workers disdain as being unworthy of serious attention. Yet what I see being produced on the higher levels of theoretical analysis of the educational process, I must say, does not impress me greatly. As long as you are working within the model, you can have all sorts of research fun, and do all sorts of elaborate things. But it's hard to show how this has any actual impact or how it ought to have any impact in the actual processes of education.

What major barriers have you faced in using the longitudinal approach?

I haven't used it appreciably so I haven't myself, faced these barriers. But I do know pretty well what they are from the reports of other individuals. There is, first of all, the fact that if you want to do a longitudinal research study of any magnitude, you've got to commit a sizeable portion of your life to getting it done. It is also likely to be rather costly. People move around. The sample that you start with tends to melt away. You have problems in keeping track of individuals. When you do track them down, their circumstances

have changed so that they're no longer really appropriate subjects for your investigation. So longitudinal research is costly and time consuming. Anyone who has engaged in such a study can tell how frustrating it can be.

What is the general promise of longitudinal research? I think it is a technique of some utility, but of definitely limited utility. Perhaps this could be said in general, about any technique. I don't agree with those who say that there'll be a great getting-up morning in educational research when we come to realize the importance of longitudinal studies, that the scales will fall from our eyes and that we will then really begin to understand human behavior.

In the fairly large university I serve, there are literally thousands of staff members involved in the educational process. All of them are concerned, directly and indirectly, with conditions that will facilitate educational growth. The complexity and the diversity of these things almost defies description. It would be nice if the educational process were simple and pure like nuclear fission which, I realize, is not really very simple. But if it were that kind of a process, once you understood its applications. I don't believe the educational process is that kind of a natural

process at all. The essence of it is variety, diversity, and complexity. Hence, I'm not at all sanguine about the amount of improvement in the educational process we should anticipate from a longitudinal study of the conditions which seem to facilitate educational growth.

To sum up, I do not dismiss the longitudinal technique as worthless. I do recognize its difficulties, and its special and fairly limited applicability. If you push me to suggest an area in which longitudinal study is urgently needed now, I'd have a hard time giving a convincing answer. Perhaps this note of skepticism or caution about longitudinal research studies is not entirely out of place here.

COMMENTS ON LONGITUDINAL METHODOLOGY

Jack R. Gibb
Resident Fellow
Western Behavioral Sciences Institute, La Jolla, California

One advantage of longitudinal research is that it is tied in with a larger perspective in looking at variables. The crucial experiment, with the clear isolation of one variable, is particularly well suited for some kinds of problems. Increasingly, it becomes clear that, in the devastatingly complex play of variables in a social situation, we need to look at the variables in action and in interaction. Many statistical methods are increasingly useful for this test of interplay. The many variations on analysis of variance handle this problem pretty well. Other problems are not handled as well.

With analysis of variance we have a test of cross-sectional interplay of variables. We have no such adequate test of temporal interplay on a multi-dimensional grid.

For instance, if we are to look at the effects of punishment upon productivity, learning, and other variables, we can easily see that punishment increases learning speed, short term productivity, and behavior change in children. The residual effects of punishment upon the intra-psychic

or intro-organismic conditions are presumably very complex, often differ from manifest behavioral effects, change over a long period, and accumulate in not-easily predictable ways. That is, early effects of punishment may be great to the teacher, supervisor, or parent. Residual effects may be very devastating upon the personality of the child or adult. The negative effects may not appear for a long time and often are not easy to associate with longitudinal "causes". The only satisfactory way to handle this problem is for us to do longitudinal research in field conditions -- good field experiments with long-term schools, families, or factories.

The same things could be said about such socially significant variables as: the effects of piece work in industry, persuasion methods in advertising, merit pay in family or work situations, sexual abstinences, masturbation, and tight discipline. In fact, most of the significant problems we deal with are subject to this kind of interpretation. Another assumption that I make is that temporal effects (intro-psychic residual effects, degeneration of influence, dissipation, adaptation to inner and outer stimuli, etc.) are not necessarily linear and that the "interactions" among these variables are not easily predictable by present equations.

The interactions are such that at the present state of our knowledge of behavior we need to use free-field longitudinal situations to gain an understanding of these interactions.

One problem is that latitudinal research in political systems, family life, and other complex interactional situations has shown clear relationships that are often reversed in longitudinal patterns. This leads to conflicting and nonsensical statements about many of the crucial variables in social-system theory (e.g., the effects of competition on mental health). It is probably true, for instance, that competition leads to increased productivity in the short-term but to decrements in productivity in the long-term, particularly when the managerial methods used in inducing competition are taken into account. Persuasion methods are another case in point--the multitudes of studies on persuasion (Hovland, etc.) show some ludicrous results when the long-range effects are taken into account. The problem is that the users of research--the practitioners--are interested in longitudinal effects not in short-term effects, except occasionally.

I know very little in the technical sense about the recent statistical methods for longitudinal analysis. Others can give more competent comments. There are some helpful

developments, however,

Barriers: I would admit that much of the longitudinal stuff has been pretty sloppy. This had led to a bias against longitudinal methods. The best people, largely, are in rigorous research, laboratory studies, learning and perception, theory, etc. The standards of data analysis and interpretation that they use are far more rigorous than the standards used by "longitudinal" people, who often come from education, child development, political science, and other "softer" disciplines.

There are exceptions--some excellent methodologists are interested in longitudinal research. I think that it is possible to deal directly with system building, theory construction, and generalization at a pretty rigorous level in longitudinal analysis. The premise is great--particularly when prediction and control are involved. Look how stupid we look when making long-range predictions (e.g., the stock market, weather, gerontology, heart disease, symptom incidence in psychosis, etc.), but these problems of long-range prediction are the real and significant ones. We have to know long-range effects in education in order to participate effectively in educational change. The short-range effects

as indicated by such measures learning speed and knowledge acquisition, are meaningless in education evaluation. We need to know long-range effects of educational procedures on such factors as personality change, family memberships, character formation in later years, and citizenship.

SOME SALIENT CHARACTERISTICS OF LONGITUDINAL RESEARCH

Harry F. Harlow, M. D.
Regional Primate Research Center
University of Wisconsin at Madison

As far as I am concerned, the salient characteristic of longitudinal research is the accumulation of repetitive measures which are simple enough to prevent one from eventually becoming swamped with a large mass of data which cannot be analyzed.

I know of no unique or indispensable theoretical assumptions other than the fact that the developmental questions can only be solved by developmental data.

The only basic assumption we have made is that there are kinds of early experience which will have profound, prolonged effects upon the subsequent behavior of the organisms. In our own work on the affectional systems, we have assumed that there are a minimum of five basic affectional systems and that failure of the early affectional systems to develop appropriately has unfortunate influences upon the affectional systems which develop subsequently.

Because of the small number of subjects which we commonly have in any one group, we are usually driven to using nonparametric measures. However, in our own research

we have run across no unusual statistical problems, and I am personally not particularly interested in mathematical models.

In our own relatively simple, straightforward studies, our researches have not been limited by problems of statistical techniques. I am certain that the very different statistical problems are encountered when one engages in long-term large sampling in human researches.

I personally do not think, and probably do not believe, in scientific hierarchial levels. Our personal approach has simply been to define a classified field as best we can, such as the maturation of learning in the monkey or the maturation of the affectional systems, and then carry out a family of experiments each designed to answer a specific problem within the total field.

The major barriers which we have faced, and are facing in our longitudinal studies, are probably specific to our own problems. One of our problems has been the development of relatively normal social groups which can be maintained as relatively normal social groups for long periods of time. Another problem with which we probably struggle is keeping the obtained data to a workable,

minimal level. One of our answers to this problem is to construct experimental groups of minimal size and replicate if we subsequently run into statistical problems.

It is my personal opinion that in many research areas the cross-sectional studies have been worked and reworked until they have been reduced to parametrica. Thus, I am convinced that many of the most important remaining problems are, of a definition, problems of longitudinal research. For this reason, it is my general opinion that many of the problems of greatest future interest will be longitudinal research problems.

THE STUDY OF SCHOOL PROGRAM CHANGES OVER TIME

Glen Heathers
Professor of Educational Psychology
New York University, Washington Square, N. Y.

I'm going to have to answer your request for comments on longitudinal research in education without giving it the time and care that it deserves. I'm caught with several deadlines would around my throat so I must answer you off-the-cuff.

You are probably aware that considerations of longitudinal research methodology have been fairly well explored in child development research. Have you seen the sections on longitudinal research methodology in Mussen's handbook on child development (chapters by Baldwin and Kessen)? In my work on age changes in emotional dependence and independence at the Fels Research Institute, I found it more practical to do cross-sectional studies (comparing children of different age groups) than longitudinal studies.

All of the difficulties listed by Baldwin and Kessen stood in the way of my doing longitudinal research, even though the Fels population was established and maintained primarily to conduct such research. In consequence, only a few of the studies at Fels in the area of psychological or

social development have been longitudinal. Key difficulties hampering longitudinal research were these: maintaining a research staff for a project over a period of years without losing key workers; controlling or measuring variables that operate over a span of years to effect the individual's development; determining an adequate research design at the outset of the study (since changing the design during the course of the study to introduce new variables or improved measures would destroy pre-post comparisons.)

My comments refer particularly to the 1958-63 study of the Dual Progress Plan in grades 3-6 at Long Beach and Ossining, both in New York. While the tryout of the DPP covered a five-year period, it was not conducted as a longitudinal study. Instead, the research design called for comparing achievement test scores of pupils in grades 3-6 in 1958 with scores of pupils who were in those grades in 1963.

Why take five years for the tryout of the plan? The main reason was because it takes years for a school system, even with extensive help from outside, to install a new organizational plan and get it working at the point of instruction. The structural features of a plan such as the DPP can be implemented fairly readily. With the DPP, these structural

requirements were establishing a semi-departmentalized schedule, making differential groupings of pupils from subject to subject, assigning all teachers of grades 3-6 as full-time specialists of a curricular area, and setting up each classroom as a laboratory equipped for teaching one curricular area.

The tough tasks in implementing any organizational plan in schools have to do with adapting the fundamentals of instruction to the purposes of the innovation. These fundamentals have to do with curricular units and sequences (as called for in nongraded instruction), measures of pupils' achievement of curricular objectives, procedures and instruments for diagnosing pupils' characteristics as learners, and teachers' competencies in conducting instruction according to the purposes of the new plan.

In the case of the tryout of the DPP, it took several years for the school systems to develop non-grade-level curricular sequences in mathematics and science as called for by the plan. It took years for the schools to work out the required new procedures and instruments for testing and reporting pupils' progress. The in-service education of teachers so that they could meet the requirements of the plan for individualizing instruction through non-grading, etc. was the toughest problem in implementing the plan. Even after five years and at the

end of the study, the in-service education programs conducted by the school systems had only gone a part of the way toward preparing teachers to teach as the DPP called for. These facts indicate the need for longitudinal research on the change processes in a school system that are involved in implementing a new organizational plan to the point where it is ready for testing (i.e., for measuring its outcomes).

A second reason for a five-year study of the DPP was that this might allow time for the effects of the new plan on pupils' learning or adjustment to accumulate to the level where they became evident. It may well be that there are cumulative effects of such instructional practices as non-graded grouping and advancement. If so, such effects would be more pronounced if a pupil had experienced such practices for several years than for just one year. (The fact that the results of our study gave little evidence for such a cumulative effect from additional years under the plan suggests that the DPP had relatively little influence on pupils' learning.)

As Julian Stanley points out in his chapter in the Gage Handbook, a control group is needed with any pre-post study to deal with factors operating during the period of the study that could affect outcomes. Note that our choice of a

pre-post design was with the purpose of controlling variables having to do with community, school system, school staff, and pupil populations. We knew these variables would not be adequately controlled by this approach but we believed that we would more nearly approximate adequate controls this way than if we used comparison schools operating under the "self-contained classroom" during the period of the study.

Communities, school systems, pupil populations, school leaders, and teachers vary so much that adequate controls for such complexes of variables would have required perhaps 50 school systems operating under DPP and 50 under the self-contained classroom. This would not have been possible with the means available to us. It would not have been sensible, either, since pilot projects on the DPP had not been conducted to work the bugs out of the plan, to work out procedures for implementing it, and to develop instruments and procedures for measuring its implementation and its outcomes.

I do not believe that longitudinal research poses any unique problems of research design or statistical method. Note that any study of the impact of conditions on individuals (as in a learning study) is longitudinal in the sense that the study measures stimulus and response events with an individual that cover a period of time during which other variables

might intrude to affect outcomes. In studies such as the five-year tryout of the DPP, such problems of control are enormously magnified.

My notions with respect to field tests of innovations have less to do with longitudinal methodology than they have to do with general problems of designing and executing projects that introduce changes into systems. Many aspects of a school system's "behavior" change when an innovation is installed. Very often, the innovation that actually is put into effect is a considerable modification of the innovation on paper; system factors in the local setting enforce some sort of adaptation of the innovation to fit the local system.

Here are some of the more sticky problems in designing and conducting field tests. (1) Ordinarily the change program involves introducing a set or sub-system of new variables into the school system. These variables interact in complex ways with each other as well as with variables already operative in the school system. To date, we lack means of ensuring that the variables involved in the innovation all are put effectively into action in the local program. Also, we lack means of measuring separately the effects of different variables (or features) of the innovation. For example, in our test of the

DPP, we could not determine the relative effects of specialist teaching, departmental organization, differential grouping from subject to subject, ability grouping within subject, etc. (2) Sponsoring agencies (such as private foundations) often demand that the study begin with a full-blown, extensive field test of the innovation--particularly if it is a new organizational plan such as the DPP. The immediate result of requiring the introduction of the change program in a number of schools (we had 9 schools in two school systems from the outset) is that the research personnel is spread so thin that nothing can be done well. Also, going into research orbit from ground zero in a matter of weeks means that there is not time to do the exploratory work required for developing a good study design, nor to do the developmental work on such matters as curriculum, tests, and teacher education that are critical factors in the success of the undertaking. (3) School systems are ill-equipped by inclination or resources to help researchers establish and maintain the conditions that are required to conduct sound research studies. The researchers must deal with almost weekly exigencies that threaten the integrity of the research study. (4) During these times (and, perhaps, at any time) one cannot conduct a study in a school system covering one or more years during which time the only major changes

introduced in the instructional program are those involved in the study. For example, both Long Beach and Ossining felt it essential that they make radical changes in the curricula used in mathematics and science. This meant that a comparison of achievement test results in these subjects obtained 1958 (prior to the DPP) and in 1963 (after the DPP had been in the school system five years) might be measuring chiefly the effects of introducing new curricular materials rather than the effects of the DPP.

I'd like to make two points that seem to me to be important for those interested in studies of change programs in schools. One is that almost all such projects fail to include in the research design provisions for measuring the actual implementation of the changes being introduced. In other words, no measures are taken of the independent variables as they operate in the schools; rather, the independent variables on paper are related to the dependent variables as measured in action.

The second point is that it is extraordinarily difficult to set up appropriate control groups with respect to change programs in school systems. The number of factors requiring control is very large. Most of such variables

cannot be measured adequately. Control by randomization would call for a large number of school systems under each of experimental and control conditions. It seems to me preferable to start with pilot studies of an innovative program, using a pre-post design (perhaps with comparison schools not employing the innovation as a way of checking on the possibility that social changes over the period of the study were affecting outcomes). Following the pilot studies, if the innovation appeared to be promising, it would be time to set up major controlled studies involving a large number of schools or school systems.

LONGITUDINAL INQUIRY: SOME REFLECTIONS

Joseph A. Johnston, M. D.
Chairman, Department of Pediatrics
Henry Ford Hospital, Detroit, Michigan

Longitudinal studies, projected into the future, have as their objective the definition of the dynamics of the processes of growth and development as opposed to cross-section studies in which one expects to describe various items that characterize the organism at a point in time.

Our own experience with a longitudinal study was most rewarding in attempting to gain information on the question of whether the adult form of tuberculosis arose endogenously from the childhood form or represented a reinfection from without superimposed on a now sensitized tissue. This was possible when we were able to follow a group of nine hundred children, known to harbor evidence of the first infection, and removed from the source where this has been acquired, into other homes free from tuberculosis. Following this group for twenty-five years we watched develop the adult form, chiefly as the children went through puberty.

The occurrence of the adult form of the disease at this age could, of course, be shown in cross-section studies of a large population at each age level but we would have learned nothing about our major objective. As so often happens,

the yield of information on tuberculosis proved to be less important than a great deal of data we were able to acquire on the metabolic and biochemical changes that characterize adolescence, these studies being undertaken in an attempt to understand how the phenomena of sexual maturation were related to the breakdown of the primary tuberculosis infection. Many other items such as the food requirement of this age group, the effect of hormones on nutrition, and in turn, the effect of this on the disease were investigated fruitfully as items related to the major objective. It should be a rule of a longitudinal study that there should be no hesitation in leaving the main road to go down side alleys.

Our other experience with the longitudinal approach has been in connection with an attempt to explore the question of school readiness. Here it was our plan to make simultaneous measurements of mental and physical progress, to observe the correlation, if any, between them, to make predictions about the child's ultimate attainment, and to check this yearly by observing scholastic achievement.

I am not sure after twelve years of this that we are gaining information from the longitudinal approach that would not have been obtained from cross-section sampling of various age groups. Our conclusions appear to be largely negative.

We have not found good correlations between physical measurements and readiness, mental potential or achievement. But again we have fresh information on the early appearance in the urine of sex hormones and of the correlation of this with bone age, though this latter has long been known, but not utilizing current improved steroid analytic techniques.

The great disadvantage of the longitudinal study is usually the difficulty in keeping intact the group to be followed. This proved no problem in the private school, but the public school group constantly disappeared with removal.

Another disadvantage is the inevitable change in the personnel of the investigators.

MY REFLECTIONS IN RESPECT TO THE
APPLICATION OF LONGITUDINAL METHODS
TO THE EDUCATIONAL PROCESS

Eric Lunzer
Professor, Department of Education
Manchester University, England

Let me say at the outset that I am speaking to you under false pretenses. For although I have devoted many years to the study of development and learning, and particularly to the intellectual processes, I have never myself carried out a longitudinal inquiry. Yet, it would be churlish on my part not to offer my comments, for what they are worth, and you for your part are free as from this moment to stop the machine and erase the rest of this tape. It were no more than justice and I should not feel slighted.

I would like to group my remarks under two headings: general considerations and technical considerations. By general considerations, I mean the uses and limitations of longitudinal methods as opposed to experimental, cross-sectional or clinical methods. I believe that clarity about the sorts of problems that can best be answered by longitudinal methods is an essential prerequisite for decision-making about such questions as sampling, statistical techniques and so on.

Now the first thing I want to say is that in any area of inquiry, the use of a longitudinal method imposes restrictions in that it entails greater structuralization from the outset. If we are engaged in an experimental study, or a cross-section normative inquiry, or a series of clinical probes, we can take several bites at the cherry. We may make a false start with one experimental group and correct our errors with another. If the study is longitudinal, we are much less free to do this. The sample is predetermined, studying the development of spatial frames of reference in children and the method is longitudinal. It is of no use deciding half-way through that the techniques we are using are unsuitable because they are too language-loaded or that we ought to have collected information about the child's freedom of movement at various stages of development and have not done so. If the method is not longitudinal we can rapidly correct our error. In calculating the pay-off matrix of our mode of study, we need to bear in mind that the input of expenditure in terms of initial planning, waiting time and organizational hurdles is very large as compared with other methods. We must, therefore, be reasonably certain that the output in terms of solid results, integrated within a frame of reference so that they can be generalized, will be commensurate. In particular, we want to

be sure, first that our planning has been such that such results can be anticipated, and second, that they are likely to be such as could not have been realized by some other method.

Therefore, one's first advice to anyone contemplating a longitudinal inquiry is don't, or rather, don't until you are satisfied as to just what it is you wish to find out and how you propose to discover it. And you cannot be satisfied as to these two points unless there has been adequate exploration by other methods. Moreover, if what you wish to establish could equally well be established by other methods, then longitudinal methods will generally be contra-indicated.

I think it is important to say this because I believe there is sometimes misconceptions about longitudinal methods. The first is that longitudinal methods constitute a panacea. They provide the answer. The argument runs something like this: We wish to discover more about development; development is process that occurs through time; therefore, if we observe that process through time we will discover what there is to be discovered.

Now, of course, this is nonsense. But it is instructive to see just why it is nonsense. It is nonsense because we cannot observe development. The most that we can do is to take a number of soundings at strategic points in the two

dimensions of behavioral content and time. Hence, the two questions: just what is it that we are going to observe? and how frequently are we going to observe it? Neither of these questions can be answered except in a specific context and neither can be answered except in a framework of theory. Any investigation if it is to be fruitful must set out from more or less specific hypotheses. Because of the long-range planning entailed in longitudinal studies and the consequent need for advance decisions to be taken with respect to techniques, the hypotheses need to be more specific rather than less so.

The second misconception is that in the final analysis only longitudinal methods can yield valid data concerning development. And this, too, is nonsense. For instance, even if not a single test of intellectual development had ever been repeated over an interval of time with the same subject, most of the facts as we know them in this field would continue to command our confidence. And they would do so by virtue of the experimental and cross-sectional studies alone. For it can hardly be an accident that all our older subjects tend to be abler than all those younger. Of course, this does not mean that longitudinal studies have no value. For instance, only longitudinal studies can provide information with respect to the shape of growth curves, and longitudinal studies are

infinitely superior to retrospective studies in regard to questions of prognosis.

Let me take two examples to illustrate what I mean with regard to the specificity of hypotheses in the planning of a longitudinal investigation. Let us suppose that we wish to discover more about the relation between delinquency and the withdrawal of love. We think, rightly, that a longitudinal inquiry would be valuable in this area. Now it is not sufficient to obtain measures of affective relations in the home in early childhood and of delinquent incidence in adolescence. For this can only tell us that there is a relation between the two, and this we already know. We need to know more about the mechanism of this relation. And this means that we have to take more measures of behavior and doubtless of affective relations along the curve of their development. Nor can the measures be simply measures of delinquency, since we expect this to be less common in pre-adolescence, and anyhow, we want to establish something about its antecedents in behavior, going right back to the child's responsiveness to his early environment. What sort of measures shall we take? At once we see the need for an adequate theory. Shall we take behavioral or test methods of extra-punitiveness and of intropunitiveness of the effectiveness of social relations

between the child and his peers, of the strength of affective ties which he forges and so on? How often do we take such measures? How do we obtain them? How do we score them or categorize them? Is our list exhaustive? If not, is it sufficient? All these are decisions which we need to make at the beginning and not halfway through. If we are uncertain as to the validity of our measures or their categorization, we have to establish that first, and the method must be experimental or cross-sectional or clinical or some combination of these. The longitudinal study itself is the final stage, not the initial.

My second example is in the intellectual field. We already know a great deal about growth curves with respect to measures of ability from longitudinal studies which are now classical. But the work of Piaget had given us a completely new insight into the kind of organization of experience that is involved in problem solving and intellectual behavior, into the nature of intelligence (or at any rate of intelligent behavior) as opposed to its measurement. And this poses new problems. In particular, Piaget suggests that the coordinations involved in intellectual behavior are of two main sorts: primary coordinations effecting concrete reasoning, and secondary coordination effecting formal reasoning. Such coordinations

are assumed to be very far-reaching in their effects. They transform large areas of behavior and produce new levels of equilibrium in the interactions between the child and his environment. Therefore, one would suppose that as these coordinations are achieved, their effects should be manifest in many fields. Since the coordinations in various behaviors are essentially similar, even though we are still unclear as to the mechanism of their achievement, we are entitled to expect a great deal of transfer to related and even to unrelated fields once a given level of coordination has been achieved in a particular area of development. For instance, there may be individual variation in the age at which children achieve the coordinations involved in the conservation of substances, of area, of number, etc., but if Piaget's analysis is correct, we would assume that when a child has achieved one of these there should be rapid transfer to the remainder. If we do not find this, then the theory needs to be modified and amplified to meet these new results. Here then is a field for longitudinal study. For though we know that some children may be successful in one area and not in another, without a longitudinal study we can have little knowledge of the degree of transfer, for here we should need several measures over a period of time. Even the work of such investigators as

Beilin and Smedslud cannot be decisive in this regard, for all the indications are that short-term acceleration of systematization is less stable and less effective than what Piaget would call spontaneous elaboration.

Such a longitudinal study would build on and extend the findings of other modes of study. Again, it is a late stage in inquiry and not the beginning. And again, it is highly specific. It must be to warrant the use of the longitudinal method. Before setting out on it one needs to be clear as to what tests will be relevant, how they are to be administered and how frequently, and how they are to be scored or categorized. In the given instance, such decisions can be reached fairly effectively, precisely because so much work of theory construction and instrument testing has already been carried out by a clinical experimental method. This enables one to make advance decisions as to the administration of the inquiry and the recording and categorization of the subjects' responses. I would emphasize that the diagnostic method used by Piaget is particularly suitable for the preliminary phase of the investigation of such processes, but it is not suitable for the preliminary phase of the investigation of such processes, but it is suitable for providing the sort of specific answers that we expect from a longitudinal study.

Interviews need to be more structured to gain accurate results which will provide an adequate foundation for judgments as to growth trends and so on. But they must be structured in the light of the evidence gained in the clinical inquiry. It is not enough to have quantitative results; they must be relevant to the purposes of the investigation.

As to the question of the form that such quantified records should take, perhaps I should say that I do not think it is important that they yield a continuous scale or scales. In some cases, they will and in others they won't. What is important is that they be in the form of at least operationally defined, reliable and valid categorizations. This does not happen automatically. Whence the need for much preliminary work. Above all, the quantitative results that we obtain must be relevant to the framework and hypotheses from which we set out.

There is one other point which would arise in an investigation of this sort. That is, that the effect of repeated testing itself may affect the course of development. If we are aware of this, we can anticipate it and allow for it, for instance, by combining the longitudinal study with experimental studies, or by the differential treatment of

of sub-groups within the sample with respect to inter-test intervals. Again, the need for much thought and planning.

Finally, even in this field, it is by no means true that because fruitful results may be hoped for from a well-planned longitudinal study, that other methods of study have become irrelevant. For instance, experimental methods are essential to shed more light on the actual mechanism whereby children acquire the kinds of coordinations that Piaget describes. And more cross-sectional studies are needed to fill in gaps, for instance, in the further systematization of mathematical concepts beyond the elementary conservation of number, the role of spatial abilities, imagery and so on in the systematization of spatial relations, the difficulties encountered by children in the categorization of temporal and spatial relations that they encounter in social studies, and so on. All these are questions that we are looking into, without touching on longitudinal inquiries yet. Longitudinal studies are useful. They constitute one method among many. A careful longitudinal inquiry is not better than a careful study of another sort. It is different, and designed to answer different questions. A poorly thought-out longitudinal study is as wasted as any other poorly thought-out study and probably more costly in terms of time and money.

Now let me come to questions of technique. Here I will be very brief, for you will doubtless have contributions from specialists infinitely more qualified than myself. I want to say a little about sampling, and just a little about statistical techniques.

About sampling, I want to say that the longitudinal study is at an advantage, in that the subjects provide their own controls. Therefore, we do not need large samples to iron out the effect of uncontrolled variables. The actual size of sample and the population that it represents must depend on the question one is examining. But longitudinal studies should not be confused with normative studies. Their purpose is never normative, if for no other reason than that the "norms" obtained are likely to be out of date. In any case, the method is too consuming of time and effort to permit working with large samples. All the more reason for care in preliminary work.

Coming into statistical techniques, all I have to offer are personal biases. I would think that the most important and useful data that one can produce are the figures of incidence themselves. Of course, one needs statistical tests of significance. But in general, these simply tell us how likely or unlikely the values that we obtain are due to change.

One can, however, go a stage further and determine the probability that the figures are out by given amounts, as when calculating the probable error of score. All this is hard and very relevant. Correlations will doubtless be used also, as they have been. But it is important to bear in mind that correlations constitute no more than a summary of the data with a great deal of loss of information; they are not a substitute. And the significance of a correlation is simply a function of the numbers involved as it is of the size of differences revealed. I am suspicious of factor analysis in the context of longitudinal studies, because I prefer to think of factor analysis as a heuristic procedure rather than a definite one. And I have rather stressed that the questions asked of a longitudinal inquiry should be rather precise. But these are no more than personal biases. Finally, when using tests like chi square and analysis of variance, it will probably be important to bear in mind that usually the data will be correlated.

I need hardly add that the length of the inquiry, like the frequency of observations or tests will inevitably be determined by the questions it is designed to resolve. Nor need I remind you that it is often possible to retain many of the advantages of the longitudinal method with a great

reduction of time by using short-term overlapping studies instead of a single long term study covering--say a 20-year span.

Finally, a word about the relationship of longitudinal inquiry to the scientific hierarchical levels of: classify-describe, understand-explain, generalize-theorize, and predict-control. It seems to me that at any stage of scientific investigation we are bound to categorize information, we are inevitably imposing some theoretical framework which determines our categories, our findings are aimed at least at a partial explanation or elucidation, in the sense of validation and modification of theory, and we are always making predictions and testing whether such predictions are sound. All I would say is that longitudinal studies have least to contribute to categorization. The necessary categorization should have been done before. They are concerned with all of the remaining three, but to each their contribution is specific. That is to say that longitudinal studies are undertaken to supply information that cannot be obtained by other means.

SUBJECTS AS A RANDOM VARIATE IN LONGITUDINAL INQUIRY

Frederic L. Matthews
Statistical Consultant-Statco
Williston Park, Long Island, New York

"Longitudinal methods in research on children" has a lilting sound. My interpretation of this phrase is that this is a method of analyzing the data of the same children by a sequential formulation in time. A group of children are tested in one year. Uniform tests are administered to the same children the next year. Uniform tests are administered to the same children sequentially for many years. The accumulation of the data of the same children for several years is a longitudinal study. It seems to me that the longitudinal method is a method of avoiding the real and difficult problems associated with the matched pairs method. The matched pairs method, as far as my own feeling goes, should be avoided like the plague. The longitudinal method, by pivoting attention on the same children, may avoid the hidden biases in selection; but is it really a step forward or backward? I know it appears reasonable; so were the good old days when researchers used to hold every variable constant but one. Research has moved a long way from that approach, culminated by the multivariate analysis; i.e., all variates are examined including the children. The

absolute condition that the children in a longitudinal study must be the same, seems to me, to be a retreat from multivariate analysis. Is there some magical or mystical value so that children cannot be treated as a random variate? The supposition that a study is simplified by using the same children could lead to serious complications.

For instance, suppose we are contemplating a longitudinal study starting with a grade in elementary school and continuing for several years. Initially, there will be N pupils, the enrollment in that grade. For easy calculation, we will assume that the enrollment in future years will be stationary. The stationary enrollment does not necessarily mean that in subsequent years, the same number of pupils will remain in the study. All that a stationary enrollment means is that the number of children in the in-migration group equals the number of children in the out-migration group. Moreover, at the beginning of the study and in subsequent years, some of the pupils will have missing scores because of illness or other reasons. The pupils in the in-migration group are obviously not in the original study, and the number of pupils in the out-migration group plus the number of pupils with missing scores is the loss of pupils from the original study.

Let us assume that the sum of the out-migration children and the children with missing scores is constant each year. That is to say the total loss of pupils in the study is proportionally the same each year. Call this proportional loss of children from the study, "a". At any time the number of children lost from the study is $-aN$. In differential form the loss is then:

$$\frac{d}{dt} (N) = -aA$$

where t is in years and a is the proportional loss of pupils. This equation is the unimolecular equation in chemistry and physics, or the radioactive decay curve. The integrated form is:

$$2.303 \log_{10} N = -at + c$$

where c is the constant of integration.

At the start of the study, the time is zero. Then:

$$+0=0$$

$$-a+0=0$$

$$N=N_0$$

The constant of integration is then:

$$c = 2.303 \log_{10} \left[N_0 \right]$$

The final equation, after substituting and simplifying is:

$$2.303 \log_{10} \left[\frac{N}{N_0} \right] = -at$$

This equation will tell at any time, t , the number of pupils, N , still in the study.

The annual proportional loss of the children can be determined in this study or in any study. In many districts or schools, the annual loss is 20%. Some districts or schools have an annual loss of 10%. A very few schools have annual loss as low as 5%.

A useful calculation is a determination of the time when the original group will have been reduced by one-half. That is to say, the half time is the time when half the students are still in the study.

$$\frac{N}{N_0} = \frac{1}{2}$$

A table has been constructed for the half time for a series of different annual proportional losses.

<u>Annual Proportional Loss</u>	<u>Half Time in Years</u>
<u>a</u>	<u>+½</u>
0.25	2.8
0.20	3.5
0.15	4.6
0.10	6.9
0.05	13.9

In the study, the teacher's instruction is a teacher-class couple. It seems foolhardy to assume that the pupils remaining in the study are exactly the same as those not in the study. If the pupils not in the study (in-migration and pupils with missing scores) are different from the pupils in the study, either higher or lower, the different composition of the entire group will inevitably affect the teacher's instruction. If the loss of students in the study is measurable, then an exact supplementary analysis must be included to justify continuation of the study. The students remaining in the study and the students not in the study must be proved comparable.

Looking from the other point of view, consider a three year longitudinal study with 95% or 90% of the original group of children still retained in the study after three years. That is to say, the total loss of children from the original group will be 5% or 10%. The annual proportional loss of children in the study from out-migration and missing scores would be

<u>Three Year Loss Study</u>	<u>Annual Proportional Loss</u>
95% retained	0.017
90% retained	0.035

Even the highest annual proportional loss of 3.5% means that in a classroom of 30 pupils, there will be only one pupil lost each year because of out-migration or missing scores. This is not realistic.

In this office, just before I started to write this note, I finished a study in one grade in a school district. This analysis is not connected to a longitudinal study. However, the number of pupils involved and the proportional loss are germane. There were 400 pupils in this grade. The print-out of the digital computer showed that 38 pupils had at least one score missing. In addition, there were 25 pupils listed in the original group who were missing the following year--out-migration. The total of missing scores and out-migration was 63 pupils. The proportional loss in one year in the study was 0.1575. Assuming that the proportional loss is constant, a three year longitudinal study would net 249 pupils from the original 400 pupils. If the enrollment is also constant, the number of pupils who are physically present but not in the study at the end of the year would be 151. Q.E.D.

The longitudinal studies that I have done fall into two types: retrospective analysis and prospective analysis.

Retrospective analysis is data in being at the present time. Prospective analysis includes not only the data in being but also the design and analysis of future data to be collected. Actually, of the two types, prospective analysis is truly a longitudinal study. Retrospective analysis is less costly. Most of the calculations in this office have been retrospective analysis.

Retrospective analysis consists of a series of functions time ordered, and could be a time ordered multivariate analysis or could be the relationship of a criterion at the present to a series of functions or relationships time ordered in the past. This analysis depends on the past scores in the school district. Furthermore, the time order is only partly ordered. This retrospective analysis considers only the pupils at the present time related to the past. It does not consider the other pupils who were there at the beginning and are then lost, nor those who came in after the beginning. Generalization and prediction are necessarily narrow if this method is done for just one year. However, the formulation of the retrospective analysis continued and validated in subsequent years approaches a generalization rapidly. Alternately, if there is data

for three years and if all of the appropriate functional relationships are the same, then a test of consistency permits a generalization or a firm prediction. Actually, my practice has been to develop the functional relationships and provide a method to validate these relationships annually. Each year the relationships are validated, the prediction becomes stronger. When the relationships are found to be no longer valid, then recalculation is necessary.

Prospective analysis is really the most rewarding type of longitudinal study. I have had one complete prospective analysis contract which is now in its second year. Several proposals of this type are now pending. The time and cost are appreciable, but the practical applications in education will be tremendous. Even the term "longitudinal", which implies the same children, is tacitly eliminated. Essentially, this type, prospective analysis is a generalization of the children which returns to multi-variate analysis.

For example, let us consider an exact longitudinal prospective analysis covering three grades. The base lines of this study will consist of the formulation in each of the three grades of the relationship of different criteria to one or more covariates. The developed formulations will be

from the present data plus as many previous years (usually two or three) to obtain stability of the formulation over several years and the generalization of the children, also over several years. The stability in the formulation at one grade and the generalization of the children (so they can be treated as a random variate) is the keystone of the advantage of this type of prospective analysis. Having developed base lines, all future classes in these grades can be compared by assuming that they are the same as the base line and testing for consistency to the assumption--or to a generalized regression. In all cases, however, the children are not constant but are generalized as a random variable.

1. What are the salient characteristics of longitudinal research?

Longitudinal research is a time sequence in multivariate analysis.

2. and 3. What important theoretical assumptions are unique or indispensable to the longitudinal approach and the role of these assumptions?

The tests in a longitudinal study must be stable or uniform. I hold that the usual restriction of the same children over a period of time inhibits the generalizations of the findings. Only by feeling the children as a random variate can the longitudinal approach be most useful.

Whatever the mathematical model employed in the longitudinal study (linear regression, covariance analysis, etc.), the actual data must be checked for consistency to the underlying assumptions of the mathematical model (normal distribution, equal variances, linearity, etc.). If the consistency of the data to the underlying assumptions has not been checked, then any inferences drawn are suspect.

4. and 5. What statistical or mathematical models are most helpful in employing the longitudinal approach? Does the longitudinal approach limit the use of certain common statistical techniques?

The longitudinal approach lends itself to any mathematical model except the very crude statistics; "t" tests between means, particularly when the distributions of the variates have not been checked for consistency to the theoretical distribution, is an example of inapplicable crude statistics. In a longitudinal study, there should be at least one or more covariates ancillary to the model so that the formulations and the children are comparable.

6. Where can longitudinal studies make the greatest contribution?

Properly executed longitudinal models using multivariate analysis with the appropriate tests of consistency of the data to the theoretical model permit the study to contribute to all four classifications.

7. Major barriers:

Cost and time of proper analysis are the major barriers. A longitudinal study necessarily freezes the tests. I do not believe that many schools will tolerate the inhibition for more than three years.

8. General promise of longitudinal research:

The promise of greater understanding of many facets of education is excellent, but the cost restricts the number of schools which will employ the method.

9. A coda.

There are very few "quick fixes" in education. Longitudinal studies need stamina on the part of the educator and the analyst both in the conception and in the execution of the long range study implied by the titled Longitudinal Studies.

THE LONGITUDINAL APPROACH TO RESEARCH

WITH CHILDREN: A DISCUSSION

Harold Mitzel
William Rabinowitz
Robert Lathrop

Professors of Educational Psychology
College of Education, The Pennsylvania State University
University Park, Pennsylvania

The following comments represent the substance of a discussion held by three professors of educational psychology at The Pennsylvania State University concerning longitudinal approaches to educational research. The discussion by Harold Mitzel, William Rabinowitz, and Robert Lathrop grew out of a request by Dr. Gerald T. Kowitz. The finished product has been edited by all three participants.

Although the comments reported here present a number of important considerations, the editors believe that the extemporaneous manner in which they are presented provides a perspective on longitudinal research which is not available in traditional textbook treatments of this topic.

We trust that the dialogue reported here will prove not only informative but will also provide some insights into the way three experienced researchers regard this methodological problem.

1. What are some of the distinguishing characteristics of Longitudinal Research?

Rabinowitz: Well, it seems to me that in a longitudinal study the most obvious characteristic is the repeated measurement of one or more individuals over an extended period of time. The longitudinal aspect involves a time dimension and the presumption is that you can measure these individuals on some characteristic over that time dimension. It could be a physical growth characteristic like height or weight, or it could be a psychological characteristic like intelligence or ability to solve problems or digit span. You could do it with one or two people, autobiographies, for example, or you could do it with a group of individuals. I don't know that it has any other salient characteristics.

Mitzel: What you say reminds me of the fact that the word "longitudinal" has some kind of implicit value judgment built into it when you are dealing with human subjects. A longitudinal study of a cockroach might involve only a two-week span of time, but when we speak about longitudinal studies of children we ordinarily talk about something in terms of years.

Lathrop: Longitudinal studies are ordinarily set apart from the contrasting kind of study which we refer to as "cross-sectional." Cross-sectional studies are distinguished by measuring a number of individuals at one point in time and longitudinal studies conversely usually imply measuring one or more individuals on several successive occasions.

Mitzel: Let me ask a question. We have heard a lot of talk lately about retention; can we slip a word about retention studies in here? Are these the kinds of studies in which we are interested in learning criteria? Do these come within the purview of longitudinal investigations? In other words, are retention-oriented learning studies longitudinal research?

Rabinowitz: I think there isn't any hard and fast definition of what constitutes longitudinal, and what doesn't constitute longitudinal, research. I guess any study that takes more than one measure of a group of individuals at more than one point in their history could be called a longitudinal study, but typically longitudinal studies involve the examination of one or more characteristics over a fairly extended period of time (and by extended I mean more than two, or three, or four weeks) where some rather substantial changes have a chance to take place, usually over a period of years.

Lathrop: Well, would a pre-post-comparison be a longitudinal study or would it be cross-sectional?

Rabinowitz: If it involves the same individuals with the same or similar measures, then I would say it would qualify as a longitudinal study, but usually, longitudinal studies don't involve any obvious experimental treatment. Again, that is just a matter of convention, the way we use the term, but most of the time when I hear the words "longitudinal study" I think of a study in which someone has gone out and measured kids in some natural environment, whatever it might be, measured them when they are five, again when they are six, and again when they are seven, etc. They may have been selected on the basis of a special characteristic like Terman's gifted children, but they aren't usually subjected to any special experimental treatment so there isn't any pre-post-comparison possible.

Lathrop: It doesn't strike me as being a necessary restriction. One might longitudinally follow the course of two different dietary patterns, one of which would be deliberately manipulated, the other of which would not.

Mitzel: Like the tooth paste test which you see on television "43 percent fewer cavities with Crest." Is that longitudinal? That's experimental, it's manipulative.

Rabinowitz: Well, I think all you can do here is to legislate some distinctions that you are going to employ in talking about this topic. I would say that nobody has legislated this usage in such a way that all of us implicitly accept and recognize it, so you could call a lot of studies longitudinal that someone else might prefer not to call longitudinal.

Mitzel: But it does have definitional problems, doesn't it?

Rabinowitz: I would say so, very markedly. Suppose you have a longitudinal research in the sense that it involves the same individuals over a long span of time, but you don't necessarily measure the same characteristics. You might, for example, measure something about children at age three and then see if it predicts a different measure taken at age eight. To take an extreme example, suppose we measure the way in which parents rear their children during the preschool years and the way in which the children function as adolescents ten years later. Is this longitudinal research?

Lathrop: Ordinarily it is not possible to use the same measure with three-year-olds as you would expect to use with fifteen-year-olds. Intelligence tests are an example of this.

At three we use primarily manipulative and developmental kinds of items. At age fifteen we use primarily verbal items and there is no good evidence that these should necessarily represent identical abilities.

Rabinowitz: Well, you could then, I suppose, distinguish what might be called "quasi-longitudinal" studies in which you pick your subjects at some age and you ask them to retrospectively tell you about their prior history. You might, for example, take a group of college students, some of whom are having great difficulty adjusting to college and others who are not; you might try to find out something about the way that they were reared as children by sending questionnaires to their parents on child-rearing approaches and then try to relate these child-rearing tactics to the subsequent difficulties that the student did or did not experience in adjusting to college. This combines some of the advantages of longitudinal research with some of the advantages of a cross-sectional study, where you pick up your data at one point in time. It is frequently used but has a lot of problems attached to it which I think we all recognize.

Lathrop: These are what are usually called "correlation studies" and I presume that they are the basis for keeping

school records in that it is possible to go back and get some antecedent data which is helpful in explaining or predicting subsequent behavior. If this were not the case, there would probably be no point in keeping them.

Mitzel: It seems to me that the problem we are confronting may be viewed a bit easier if we think about a parallel problem in economics. Economists are always seeking a relationship between predictors and criteria as they operate through some span of time hoping that they can take their relationship and apply it to predict some future event like Gross National Product. What they have to depend on is the fact that the variables haven't changed as a result of shifting in time. It is a little easier I guess to see how calendar-oriented variables would effect the relationship in the case of economics than it is in education, but the same phenomenon would certainly apply.

2. What are some of the methodological problems in conducting longitudinal research?

Mitzel: We have been talking about some of the assumptions that are indispensable to longitudinal research and we have not been able to find anything that is indispensable to any great extent. Can we now turn our attention to the methodological problems involved in conducting longitudinal studies?

Rabinowitz: It seems to me that in any long-range longitudinal study you automatically run into the problems of securing subjects who will be able to participate for let's say twenty years. Ordinarily such subjects would be persons who are reasonably well placed socially so that they have stable addresses, they know what the research is likely to entail, they are essentially cooperative, they communicate easily, they meet their appointments, etc. Individuals who are very poor, transient, not well motivated, and mobile tend to drop out and get lost and as a consequence you are likely to find at the end of some longitudinal investigations a rather highly selected group of individuals serving as subjects.

Lathrop: There is another kind of attrition involved in a study which a graduate student here (Penn. State) was conducting, with another faculty member needless to say.

The student was interested in the development of liberal and conservative points of view over the college span of years. His procedure was to administer a liberal-conservative instrument to freshmen and then to administer it cross-sectionally to a group of seniors. He was interested in drawing some conclusion about the development of liberal-conservative points of view over the four-year college period. It seems to be completely inappropriate to make any inference about the development of liberalism or conservatism or any other developmental pattern on data which are this subject to attritional effects.

Mitzel: You spoke about attrition, I thought you were going to talk about attrition of subjects.

Lathrop: I am talking about attrition of subjects, natural selection and unnatural selection.

Mitzel: Non-natural would be a better word.

Lathrop: It may very well be in this example that students who are liberal do not remain in college, particularly in certain kinds of colleges. It may be only the conservative students who are able to tolerate certain academic conditions and remain to graduate. At least that is a possibility that exists.

Rabinowitz: If you are going to attribute any changes unambiguously to the influence of college, wouldn't you need a control group of students who are in the same age range but are not in college at all? I am saying that even if you had longitudinal data and you were going to use the data to substantiate a hypothesis about the effect of college attendance, you really would not know that it wasn't four years of experience from eighteen to twenty-two years of age.

Lathrop: That is true.

Rabinowitz: So you would need some kind of control group.

Mitzel: So what you should have would be kids eighteen to twenty-two in a culture who are not going to college as a comparison group, perhaps kids out working.

Rabinowitz: If it is really college that produces certain effects, these kids who are out working should show a distinctly different pattern of liberal versus conservative development.

Lathrop: Another problem with many of our longitudinal studies is that they are conducted over too brief a period of time. It may be that many of the treatments that we have tried and rejected in education as giving nonsignificant results are

nonsignificant because they have occurred over too short a span of time. An interaction between some treatment and individuals may actually, in the long run, fail to demonstrate a superiority or inferiority because we have taken too short a look at it.

Mitzel: Yes, but isn't that a characteristic of the treatment, the length of time it is supplied is like the treatment of fertilizer on a field. You know you can put one hundred pounds of lime to an acre. If you only do it once you will not get the effect that you would have gotten, if you did it repeatedly over a period of four years or five years. You get a cumulative effect.

Lathrop: Yes, that is exactly my point. I think that many of our studies have not given this cumulative effect a chance to manifest itself.

Rabinowitz: An illustration of that, if I can interject, is the recent interest that has been expressed in what happens to students when they go to college. Do you know the volume that P. E. Jacob produced on changing values in the American college? He came to the conclusion on the basis of a survey of the literature that by-and-large college did not significantly alter student values. He had summarized there

both cross-sectional and longitudinal research, such as it was, and it seemed to me that a study of this kind depends crucially on at least two things. One, is the assumption that the instruments that people are using in these investigations really do adequately sample the kinds of variables that can change when a student goes to college, and that is a very dubious assumption since most investigators use things like the Allport-Vernon-Lindzey Study of Values, which is a rather crude instrument. If there are subtle changes taking place, as a function of going to college, they will not necessarily be reflected in such devices.

The other thing is the assumption that if there are changes taking place among persons who attend college, they will be reflected in the comparison of his status as a freshman and his status as a senior; and yet it may be that the true implications of going to college do not really manifest themselves until one is about 30 or 35 years of age. Maybe they show up more clearly a little later.

Lathrop: Another very difficult problem in longitudinal research is the Hawthorne effect. Once a person knows that he is going to be a part of a study and follow it up, we have no

way to assume that this in some way does not affect his subsequent behavior. This is a problem which is less serious in cross-sectional studies.

Rabinowitz: The point you are making, that once people get involved in a study the very act of participating changes them markedly, is quite true. You find this in certain panel studies where we are interested in public opinion on various controversial issues, particularly where we are interested in how people make up their minds. During the past presidential campaign, for example, a number of investigators were interested in the dynamics of how people make up and change their minds, so they assembled panels and systematically interviewed these people throughout the course of the campaign. The very fact of being a part of such a panel and knowing that you are going to be interviewed week by week, it seems to me, must change the way in which you formulate your opinions.

Lathrop: There was a period in the public opinion survey field when survey groups would try to get an idealized community which they would survey very intensively and then try and extrapolate to larger populations. Of course, as soon as the community found out that it was earmarked as an idealized community it no longer could be used. I suspect that the same

phenomena holds true in a more microcosmic sense when individuals know that they are part of some longitudinal study.

Rabinowitz: There is also an economic problem in maintaining access to people over long periods of time.

Lathrop: I am convinced that long-term follow-up of people is possible if you are willing to put the economic resources into tracing people. A study done by Campbell at the University of Minnesota involved following up approximately 800 people who took the Strong Vocational Interest Blank when it was originally being developed. Through intensive effort, Campbell was able to find over 790 of these people some 25 or 30 years later. This, of course, entailed a great deal of effort and expense in tracking down a relatively small number of people who were difficult to find. One must, however, go through the effort of obtaining the longitudinal data on every possible member of the sample in order to avoid bias. Numerous investigations have shown that it is the group of people who are most difficult to obtain that represent the largest potential source of bias in a longitudinal study.

As most graduate situations operate, the longitudinal study is almost an impossibility for a thesis research. Students are not willing or able to wait the length of time to

get their data. I suppose an alternative would be for thesis advisers to develop a research program which would allow one student to begin a longitudinal study and another student subsequently to follow it up. At least, in my opinion, there are many studies which ought to be approached longitudinally which are now approached cross-sectionally because of the fact that graduate students are not willing to wait for the kind of data that they would really like to have. I think that faculty advisers could help obviate this problem by themselves maintaining the continuity which is important in longitudinal study.

Mitzel: I think that is a good idea, but there is one kind of thing that militates against it and that is the extent to which we have faddism in psychology and education. You take the area of personality tests; a guy who was, say ten years ago, interested in personality tests is now working with a dead area and few people are interacting with him, so there is an awful lot of faddism which militates against longitudinal studies in education. The situation is a consequence of a lack of good solid instruments that last and last.

Lathrop: They too have been evaluated cross-sectionally. The instruments do not seem to be useful for measuring short-term

effects and, therefore, we have abandoned them. It may be that there could be some of these procedures which could prove useful if given a reasonable span of time over which to measure the phenomenon with which we are concerned.

Rabinowitz: It seems to me that the longitudinal investigation presumes not only a group that you can investigate longitudinally, but it also presumes investigators who have a kind of longitudinality. It presumes a stable organization with people whose interest will remain reasonably firm, and who are organized to get data over a period of time and not follow fads and fashions.

Lathrop: The problem with longitudinal data is that people are not willing to make the kind of professional commitment that is required to get them. So many kinds of studies that we would like to have based on longitudinal data are never done because people are not willing to devote the kind of energy to designing the study and following it up. It is easier to do the cross-sectional studies. You can get in and get out of them but it takes someone like Terman who is willing to spend a major part of his professional life for over 35 years following up one group of students, collecting one kind of data.

Mitzel: Do either of you know how they have resolved this problem with respect to Project Talent? I know there are both cross-sectional and longitudinal aspects.

Lathrop: It is almost entirely cross-sectional. There is very little longitudinal data in it.

Rabinowitz: Yes, but they have built-in procedures, as I understand it, for collecting additional data on some of the same subjects, isn't that true? I do not know whether they are the same tests with the same people, but there are follow-ups of some of these people.

Lathrop: Do you mean retesting the same individuals?

Rabinowitz: They are mostly eleventh-graders. I am not sure whether it is retesting at all. I think there are provisions for getting more data on these people. It may not be the same data. It may be quite different data, the kind of thing "who went to college and who did not." Follow-up data is, I believe, provided for in the basic plan of Project Talent.

Lathrop: I think the development in recent years of federally sponsored research centers is going to enhance the possibility of doing longitudinal studies. One need not plan on getting small individual grants to support long periods of

study. Larger grants to centers will automatically insure the possibility of doing more longitudinal research if we can find enough consistency in personnel to do it. I see this as the major problem that is standing in the way of real implementation of longitudinal research.

Rabinowitz: It seems to me that another important function of longitudinal studies is to correct some of the thinking that inevitably follows from cross-sectional studies. Cross-sectional studies typically yield a kind of growth function or curve based upon cross sections of individuals of different ages. The longitudinal studies when compared with such cross-sectional curves frequently indicate that given individuals don't conform very closely to the cross-sectionally-produced curve, so that individuals may show quite different patterns of growth from the one suggested by cross-sectional methods. This is by virtue of the fact that some individuals grow slowly at certain points and then speed up and then grow slowly again. There are all sorts of anomalies in growth when one looks at specific individuals.

Mitzel: But isn't that a statistical artifact? The fact is that you are using the average in cross-sectional studies which tend to remove the variability?

Rabinowitz: Well then your cross-sectional studies cannot ever reveal these individual growth curves, can they? The longitudinal studies can show that there are marked differences in growth functions that may look sharp and clear over a period of years when you consider many individuals, but these cross-sectional statistics actually include some individuals who are growing far more rapidly than others and some who are not growing at all. You may really have two, three, four, five or ten distinct groups producing one curve.

Lathrop: If there are no systematic changes over time, if the group of three-year-olds now look like the group of three-year-olds years ago, it does not seem to me that this is an important problem. The real problem is that our society changes. Three-year-olds today are probably not going to look like our present ten-year-olds did seven years ago. They are going to be quite different. Curriculums are changing, societies are changing, we have changing objectives and expectations so that the use of cross-sectional data is frequently only a stop gap to the kind of data that we would really like to have. We may have to settle for cross-sectional data for the time being, but in the long run it is the longitudinal data which we will need to adjust our previous cross-sectional findings.

Mitzel: One of the questions which we ought to consider is, "What statistical and mathematical models are most helpful in employing the longitudinal approach?"

Lathrop: We have to do a great many things in the way of experimental design to compensate for the fact that we do not have longitudinal kinds of data. Many of the problems in statistical control, matching, balancing, co-variation, etc. are really attempts to over-ride the problems which are very easily and simply dealt with by using the same subject as his own control in repeated measures designs. Many of the elaborate statistical procedures are simply not required by longitudinal studies and therefore one can employ relatively simple statistical designs.

Mitzel: You pointed out that the analysis of variance where the subject is used as his own control is certainly a powerful way of approaching the problem. It is a way of getting lots of precision into a study. Are there any other ways that should be mentioned, any others that you can think of?

Lathrop: Certainly the kind of designs that the economists use, trend or time analyses, are appropriate.

Mitzel: You mean something like moving averages?

Lathrop: Yes, that would be one example.

Rabinowitz: Well there has been a flurry of interest in the whole problem of the measurement of growth. Davis' book on Educational Measurement gives a great deal of attention to this as a special problem, "how does one assess growth," and of course the book edited by Chester Harris is another manifestation of the same interest in the measurement of change, etc. There are some rather ticklish problems in this process as you know which are not very easily handled, but there are techniques for dealing with them.

Mitzel: Name some!

Rabinowitz: Well, a lot of the problems really involve trying to determine just whether or not a group of individuals measured at one point and then measured at another point have changed. That is a far simpler venture than trying to determine whether or not Jimmy has changed more from age three to age six than Johnny has changed from age three to age six. This latter kind of problem which concerns people working with individuals, like guidance counselors and others, is one that seems to me that gives the greatest amount of trouble.

Lathrop: You mean when dealing with fallible instruments you have the regression phenomena and the general problems of unreliability of difference scores so that if you are taking

an increment of change over time the difference between two points represents a highly unreliable variable. Because of the high correlation between two measures of a property such as height taken at two points in time, one ends up with an unreliable difference score. There is also the annoying problem that when you consider difference scores you alter the metrics so the difference between two I.Q.'s can no longer be regarded as an I.Q. It is a free-floating index of unknown dimensions. There are many problems in assessing growth increments which are now just coming to the fore.

Mitzel: Let me come back to a problem that we talked about a few minutes ago. I once worked on a follow-up of Ph.D. graduates at the University of Minnesota, people who had been out a certain number of years. We had no prior measures on these individuals but we knew something about them in a sociological way about their experience at Minnesota. Could we consider this kind of follow-up longitudinal research?

Rabinowitz: Were these the same people at different points in their history?

Mitzel: That is right. At different points the graduates of 1934, 1935, etc., up to 1948, or something like that. A fifteen-year span of time.

Rabinowitz: Yes, I would say so, if you have a hospitable definition of "longitudinal."

Mitzel: I was wondering, though, if that isn't too broad a definition. How could you make inferences from those kinds of data about the graduates of the next fifteen or twenty years? You certainly could not regard it as very rigorous I suppose.

Rabinowitz: You probably would have to get involved in a research effort which was longitudinal and cross-sectional simultaneously. You would get certain longitudinal data, but at the very time when you were following up one group, you would be taking cross-sections of other groups and relating both kinds of data, one to the other. If we stick to a somewhat more limited conception of longitudinal research, we probably could define it by saying that essentially longitudinal studies consist of measuring individuals on the same psychological or physical function across a period of years. If you do this, if you say that it must be the same characteristic that you are measuring and the same individuals, then you have created a problem. How do you know that it is the same characteristic that you are measuring, particularly in the realm of psychology? A physical characteristic, like height, is not so much a problem, but suppose you are interested in aggressiveness and you want

to measure it among a group of three-year-olds and then when they are four, five, and six, etc., up until the time they are 25 (if you have got the patience for it). Well, aggressiveness, however we define it, probably manifests itself in different ways at age three, age ten, or age 25. It is more likely to manifest itself in an overt and physical way at age three and a more covert, subtle verbal way at age 25. But is it the same thing then? Is an insult at age 25 the same thing as a punch in the nose at age three? That represents a rather serious kind of problem. We have a similar problem in intelligence testing as well as the measurement of social behavior. How do we know we are measuring the same thing?

Mitzel: We approached this when we talked about standardized tests, you know. We tried to develop scales and other devices that we think give us equal measures.

Lathrop: These are all developed cross-sectionally. Even if we articulate a series of achievement tests, for example, and the test given at grade three is articulated with the one given at grade four, etc., these are all developed cross-sectionally. Different populations for every age group.

Mitzel: Basically that is the way Binet developed the first intelligence examination. It was cross-sectional. There was no longitudinal approach in that.

Lathrop: Most longitudinal studies, I think, have the purpose of being normative and descriptive rather than being analytical and predictive and that seems to be the crux of the whole problem, at least as they are employed in the area of physical growth. One can set up norms over a period of years for children and then use the norms diagnostically to identify children who are deviating one way or the other from the normal pattern, but the Wetzell-Grid for plotting physical growth is an example of this in types of cumulative records which some schools keep in achievement tests. I think that whether being normative or descriptive is an essential characteristic of longitudinal study is less important than the fact that it is one of their most common characteristics.

Mitzel: In terms of the various objectives of science, classification, explanation, prediction, control, theory, I think you have offered us a judgment that longitudinal studies make the greatest contribution in helping us describe or classify.

Lathrop: I think that is the most common use for them whether or not there are other alternatives is another question. Perhaps one of you may want to address yourselves to that question.

3. What is the relative contribution of longitudinal research in educational studies?

Mitzel: It seems to me that when we look at longitudinal studies in journals and textbooks we automatically put some kind of halo around them implying that somehow or other they are better for cross-sectional data just in general, and I suspect that we have erred in this direction. That is, there is no reason to regard longitudinal data as better than cross-sectional unless it is particularly pertinent to what you want to find out.

Lathrop: I don't agree.

Rabinowitz: I am not sure that I agree either. One of the reasons that I think longitudinal investigations are valued somewhat more highly than cross-sectional studies is perhaps that it is really considerably more difficult to get the basic data, that it took more time and effort. It really is, in many respects, harder data to come by which explains why people pay a little more attention to it, rightly or wrongly. Think of the groups that have done longitudinal investigations; they typically have involved lots of people working at fairly high levels of skill over an extended period of time with fairly good resources.

Lathrop: At least in developmental areas where we are interested in growth the results of the longitudinal studies are worth examining very carefully and thoughtfully. These tend to be studies to which people go back again and again rather than the kind of study which flowers and then dies very quickly and which passes from the scene relatively unnoticed. I think that there are long-term professional pay-offs for doing longitudinal research on significant problems.

Rabinowitz: An example of that is the recent presidential address of R. L. Thorndike to the educational psychology division of the American Psychological Association. In that particular address he talked about problems in the study of growth and used as the basic data for illustrating some of the dilemmas that you get into when you measure growth, data that was collected in the Harvard Growth Study which went back several decades. He pointed out, among other things, that there was a gold mine of data there that was never analyzed because at the time there were no computers available. Today, according to Thorndike, nobody is really working with these data, but in his opinion, and I share his view, it is far more valuable data, in many respects, than that which is now being very diligently collected by doctoral students and investigators using cross-sectional procedures.

Lathrop: It seems to me that the study of learning in children would be greatly enhanced by more longitudinal studies like Piaget's and Bruner's which have access to the same children over a long period of time to follow their individual growth and cognitive development. It seems to me that this is our most useful avenue of study for understanding the learning process.

Rabinowitz: Yes, it is interesting that some of the most exciting ideas that have been generated in psychology have come from people working essentially within the longitudinal mode. I think that we can accept the notion that Freud was working with longitudinal data. By-and-large what has been so valuable in the work of such people is not so much the sophisticated methodology they used, for the methodology is recognized to be rather crude. Basically it was the way they were looking at their subjects which makes their work so valuable and so convincing.

Mitzel: I think in these last few minutes we have attached the general problem as we see longitudinal research. We have considered various questions which we feel are important and significant. If I understand what you are saying, you believe that the longitudinal method has a great deal of promise and have offered some suggestions and have

outlined some of the problems which might be considered by the person interested in this methodological approach.

SOME REFLECTIONS ON MY OWN RESEARCH AND
EXPERIENCES WITH LONGITUDINAL METHODOLOGY

Ralph Ojemann
Director of Child and Educational Psychology
Greater Cleveland, Ohio

First, I think I would like to distinguish two ways of using longitudinal studies. One can follow children over a period of time with no control over the experiences that they have. In other words, follow them as they live presently in their home, and school and community. But, there's another type or another use of the longitudinal study and that is to introduce some experimental variable, use experimental and control groups and let the experimental variable operate over a period of time. I'm not sure if this second type qualifies, but we have used it letting an experiment run over a period of two years. I suppose that that should qualify as a longitudinal study. It is the second type that we have used. We have not used the first type; namely, the type of longitudinal study in which one follows children over a period of years and doesn't pay any attention, except in a general way, to the experiences which they have, or perhaps I should say one doesn't make any attempt to introduce an experimental variable.

The trouble with both types of longitudinal studies is that if the period covered from first to last testing is

too long, it becomes difficult (it seems to me) to control the effect of other factors. And this makes generalization difficult.

If we consider longitudinal studies of the first type; namely, following children with no special control over the experience they have, then, I'm not sure that this contributes a great deal. Let's look back over some child development research which has used this method. I suppose one example is that of Piaget. The outcome of such research, it seems to me, is essentially to set norms or to attempt to get some inter-correlations. Take, for example, Piaget's study of the child's conceptualization of specific gravity. He indicates various ages at which various aspects of the concept of specific gravity develop. Actually, what he is reporting is the development that he found in children of different ages as he happened to find them. Among other generalizations he indicates that it isn't until about nine years of age that the child really has developed the concept.

We ran some studies with kindergarten and first grade children, in other words five and six year olds, in which we introduced a program for developing the concept of specific gravity and a fair proportion of children at both the kindergarten level and first grade level developed the type of

concept, or the extent of the concept of specific gravity that Piaget would say doesn't appear until nine years of age. Well, similar results have been obtained with other types of norms.

I suppose another example is found in the many numerous investigations which have been made of the birth weight of infants. And, as you know, they use these measures of infants followed over a period of time to set certain height, weight, age standards. Actually, these were simply averages of the weights of the children at the specified ages and nothing more.

The fact that they are of no greater significance is indicated by the fact that the average weights of, let's say two month old children now is quite different from the average weights of two month old children, let us say, at the turn of the century. In fact, not too long ago, one of the nutritional research organizations showed a series of graphs taking three different periods--at the turn of the century, and about 1920, and presently--that is, within recent years, and all three curves are different. Take any particular age point, say six months, and you find a different value for the weight of a six month old child. One begins to suspect immediately that such factors as nutrition and the like--that is to say our knowledge

of nutrition and its application which has changed considerably in this century is one of the important factors which has made the difference.

But the question here is, what is the advantage of what is gained from a longitudinal study of the weight of children? Well, I think that the example which I indicated suggests that one can't say anything more than that this represents the average weight of children who lived under those conditions and as soon as one begins to change the conditions, even in body weight, there is a possibility that the averages may change. It has changed as these studies in the past have shown.

I suppose the longitudinal study has a place in investigating the rate of development of something such as walking, for example, which doesn't seem to change a great deal under different conditions and in different cultures. Even there I think we ought to be very careful, because one never knows what we might learn about the development of muscular coordination and, while at the moment I can't see any particular reason why one would want children to walk earlier or later than they do now; nevertheless, from a scientific standpoint, to understand that phenomenon someday experiences may be devised which might make a change in the age of walking. I think we

would want to keep that in mind.

But when it comes to the other areas of development where changes have been made as a result of the introduction of experiences of one kind or another it seems that the norms and even the inter-correlations between different factors which have been derived from longitudinal studies without control over the experiences which the children had, have not made a great contribution.

It's quite possible that I'm overlooking something here, but in my own research, and you asked me to reflect on my own research and experiences, I found it much more productive, much more fruitful, to use the type of study in which one introduces an experimental variable, and then lets it operate over not too long a period of time. I'm willing to settle for a year, or let's say two years for an experimental variable to operate if necessary. But even then, other factors might enter in and one never knows how to take those into account when one is trying to interpret one's data. So I think what has happened to me in developing research programs is that I put more emphasis upon short-term studies--I suppose that is what you would call them--introduce various experimental variables and test for the effects of those variables.

I suppose the problem here, is the question: suppose we take the short term studies, will we ever be able to put them together so we can get the big picture of development? I rather think we will be able to do that.

Let's go back again to the concept of specific gravity. I think we could take those first grade children, for example, which we put into an experimental group and corresponding control group and where the study ran over a week, let's say, that we could have a series of weekly studies and we could trace the development of different aspects and concept. We could point to changes in the growth of the concept which took place as the result of various kinds of experiences which we introduced.

I suppose there's still the question about what kind of a development of this concept would we expect if we had experiences A, B, C, D, and E, let's say, operating over a long period of time. I haven't thought this through entirely, but it seems to me that a whole series of short-term studies with careful control of experimental variables will throw more light on what kinds of experiences will bring about what kinds of changes than the type of study where about the only variable that you have considered is that of age. And this,

it seems to me, is the case when you have longitudinal studies which make no attempt to control the experiences which the child has had.

Perhaps someone could say that in a longitudinal study, you could go to great lengths, take great pains to describe the experiences which the children are having, even though they are not especially planned. I rather think, however, that such studies as these by Barker show that this is not an easy thing to do.

To describe the ecological environment in its complete complexity is not an easy thing to do and furthermore, different children will be in different environments. What we'd have to do is work with large groups of children for whom the environment would be essentially the same. In other words, we'd have to describe enough children, wait around long enough until we got groups of children which we could then put into various groupings in terms of the kinds of environments in which they lived. This, it seems to me, would be more complex than that of taking segments and controlling experiences over a short period of time and in that short period of time, introducing the experimental variables.

WHEN LONGITUDINAL INQUIRY IS ESSENTIAL

Hanuš Papoušek

Chief, Department of Infant Physiology
Institute for the care of Mother and Child
Prague 4, Czechoslovakia

The term "longitudinal research" covers a rather wide area which can include different approaches, connected perhaps with different purposes and goals, and therefore, resulting in a variety of methods. Generally speaking, any investigation exceeding in its duration the limits of the usual one-session examination can be called longitudinal, but we usually reserve this term for examinations repeated at certain time interval in the same subjects, in contra-distinction to "vertical cross-examinations" where developmental trends are studied by comparing the means of independent samples of a population taken at different age levels, without exact evidence whether these means give a true picture of the developmental changes of the observed phenomena in individual subjects.

Even within the limits defined above, however, we can bear in mind different concepts when speaking of longitudinal research. We may, for example, use this term for short one-session tests, if we repeat them regularly during a certain age span in order to study developmental changes

in given functions. Regular yearly measuring of some anthropological parameters in the same subjects can be given as a typical example. We could call longitudinal study any simple test which must be carried out some time after some specific treatment, if it is known that the result of such a treatment can appear only after a long time interval. An example here would be a project studying the influence of early postnatal social deprivation in experimental infant animals upon their fertility in adulthood.

Or, we can also use the term "longitudinal research" when using methods which themselves represent a long term process, such as longitudinal training of a certain new ability under experimental control. Since my own research can represent a typical example of this last category, I am going to discuss some aspects of this particular problem.

Since 1954, I have studied very early development of some mental functions in human newborns and infants. This study has been part of a complex research project oriented toward understanding the development of individual differences in higher brain functions, which is still being carried out by psychologists and pediatricians working together. Among the different methods used in this project, conditioning methods are included which represent in themselves a good

example of longitudinal inquiry. I, myself, have studied appetitional conditioning.

Conditioning methods, as you may know, can include a set of several successive procedures, such as the establishment of a conditioned response, its extinction, reconditioning, conditioned discrimination and one or more reversals of conditioned discrimination. In very young infants, the application of such a complete set of methods can take several months, since individual conditioning procedures can require several tens of experimental sessions. Our project also involves simple longitudinal control of further somatic and mental development, but since these methods are less problematic, I will concentrate upon some aspects of longitudinal conditioning studies in infants during the first half-year of their life.

Let me first describe my technique. For appetitional conditioning, I chose head movements as a basic response, since head movements become coordinated relatively early, can be easily analyzed in terms of latency time and intensity, can be easily recorded with the help of an original apparatus, and can be used for comparative studies where head turning to the left or to the right are associated with different treatments.

In my experiments, the infant is trained to turn his head to the left when the sound of an electric bell is applied. Successful response is rewarded by milk presented from the left side as soon as the subject turns his head. In one experimental session a day--which substitutes for one of the scheduled feedings--ten trials are carried out. When the infant reaches the criterion of five successive positive responses out of ten trials, we start extinction of the conditioned response, i.e., we stop presenting milk and apply the sound of bell without reinforcement. When the infant stops turning his head to the left--the criterion being five negative successive responses out of ten trials in one session--we re-establish the conditioned reflex to its original level, and start conditioned discrimination.

In conditioned discrimination, two conditioning stimuli are used: the sound of the electric bell reinforced by presenting milk from the left, and the sound of an electric buzzer reinforced by milk from the right. This time, the final criterion is six successive positive responses out of ten, three to bell and three to buzzer in random order. After reaching this criterion, a reversal is studied. Now, the sound of bell signals milk from the right, and the sound of buzzer signals milk from the left. Re-reversal to the original

conditioned discrimination follows, until an analogous criterion of six successive correct responses is reached.

The number of experimental sessions; i.e., of experimental days, necessary to reach the mentioned criteria, varies in dependence upon age. That is, to establish the conditioned response takes eighteen sessions on the average during the first month of life, four or five sessions at three months, and three at five months. Conditioned discrimination represents the slowest procedure; it takes twenty-two experimental sessions in the youngest group at the age of two and one-half months, eighteen sessions at the age of three and one-half months, and seven sessions at the age of five and one-half months. On the whole, the set of all procedures takes an average of seventy-six experimental sessions, if started at the age of three days, forty-eight experimental sessions, if started at three months, and thirty sessions, if started at five months. After this basic set of procedures, several other conditioning procedures are studied, such as conditioning of affective behavior, preference studies using different tests of milk reinforcement, etc., but it would be difficult to describe all of them in detail.

Naturally, it is advantageous, if not absolutely necessary, to keep the infants at the institute during the

whole time of our investigations. That is what we really do, and what brings some general problems.

We keep our subjects at a special research unit for the first six months of life. They are usually born at our institute, and their mothers, for different social, economical or medical reasons are interested in taking advantage of this opportunity. They can stay with their infants, and get meals, linen and all necessary care free of charge. They also appreciate our interest in longitudinal follow-up study, since it is usually connected with a specialized consulting service, and mothers are invited to see our pediatricians or psychologists whenever they meet major problems in health care or up-bringing.

The main problems connected with this kind of longitudinal research are problems to do with experimental design, problems involving the basic care of the children, and economic or organizational problems.

As far as the experimental design is concerned, difficulties can appear in connection with standardization of the basic experimental conditions. It is a question of the comparability of basic life conditions among all subjects on one hand, and of the experimental conditions in relation to the long course of conditioning in individual subjects,

on the other hand. Even if it is impossible to neglect different individual demands of individual infants (and their parents), and to keep the life conditions at the research unit strictly comparable, it is still much easier to control them (keep them at least relatively comparable), and analyze them, than if the infants were brought up in different homes.

In our case, both the external environment, and the schedule of feeding, activities during walking, and sleep are relatively uniform. Mothers who cannot stay with their babies are substituted by specially trained nurses who do not take part in night shifts, and therefore, can care for infants for the greater part of their waking time.

The general state of the infants is regularly checked by pediatricians and psychologists. An analysis of the influence of several different environmental factors that could be considered potential determinants of individual difference in studied brain functions showed that, during early infancy, the appearance of interindividual differences is relatively independent, (e.g., no significant differences were found related to proportion of breast feeding as compared with artificial feeding, or to the influence of seasonal changes). In the youngest infants, the amount of total

daily caloric intake was inversely correlated with the rate of conditioning; conditioning proceeded more quickly in infants with lower caloric intake.

Besides ensuring comparability in general life conditions, we have to respect the basic requirements of comparable experimental conditions in all experimental sessions. It is well known that in longitudinal projects, slight technical improvements or even the replacement of damaged parts of experimental equipment can cumulatively lead to substantial changes in experimental conditions.

Also, the experimenter's attitude can undergo certain changes within a longer time span.

In our situation, we also meet another problem--the comparability of the general state of wakefulness and of hunger between individual experimental sessions. As you know, it is enormously difficult to have the young infants equally awake, quiet, and hungry for experiments. It took us a year of experimenting at the beginning of our project to find the best way of solving this problem. The solution was to keep the order of sleep, feeding and waking constant and rather strictly scheduled. Our infants are fed after their sleep, then they are exposed to social and emotional stimulation, and thus kept waking for an adequate time, and

again they sleep before the next feeding. Sleep in the fresh air through the whole year helps to develop a very well fixed schedule.

Speaking of the general care of the children, you must realize that we take over the responsibility for proper development of all functions in our subjects, and for preventing any symptoms of institutionalism; i.e., of social and emotional deprivation, and also for preventing contagious diseases, the risk of which can be greater in an institution population than at home. I think that it is not necessary to go into detail, but I would like to mention that while strict hygienical measures can help us prevent infection and, hence, diseases, often they are connected with a certain amount of environmental, social or emotional deprivation, or at least with a lack of variety in external stimulation (nurses in uniforms, prevailing white color and washable materials, monotonous environment and feeding, etc.). It is, therefore, important to find a reasonable compromise.

It is certainly not necessary to point out the expense of such a research unit, particularly if we realize how long this kind of investigation lasts, and how slowly one can collect such data. One way of reducing this problem is to use the subjects for as many studies as possible during their

stay. In our case, we work as a team of investigators, using the same infants for different kinds of observations, not just for economic reasons, but mainly because we want to have many sided information about different mental functions. We also want to approach the same problems from different aspects and by different investigations of the same subjects in order to be able to draw out more general conclusions on the studied problems. The number of simultaneous studies cannot, of course, exceed certain limits, particularly if the chosen methods would cause increased restrictions or physical or mental stress. According to our experiences, two or even three different conditioning methods do not cause any troubles for the infants if their application is carefully organized and scheduled. Observational methods which do not interfere with normal freedom in the infant's behavior and do not represent any physical or mental load, can be applied without any similar limits.

3 Another methodological problem worth mentioning is related to the peculiarities of early infancy, particularly to the fact that many functions undergo dramatic developmental changes during the first year of life. It is not easy to find methods for studying conditioning abilities that will be equally suitable for infants of different ages. The

conditioned eye blink reflex belongs to the category of responses that may be studied in a comparable way both in infants and in adults. On the contrary, conditioned sucking, so often used in newborns and younger infants, is an example of a method the use of which is limited only to early infancy, and therefore, is not suitable for longitudinal developmental studies.

In our particular case, conditioning methods were not used repeatedly in the same subjects; their longitudinal character refers to the relatively long course of the learning process in our subjects. If in other projects this type of method should be used for repeated investigations in the same subjects, the experimenter must keep in mind that it might be difficult to separate the effect of pretraining from the effects imposed by development, age and maturation.

The longitudinal approach can also limit the use of certain common statistical techniques. Often, and that was true in our own research, too, this approach is connected with a substantial limitation in the number of subjects in the studied sample. Thus, on one hand, a question could be raised whether parametric statistical methods are really justified. On the other hand, from the statistical point of view, the fact that a sample of identical subjects is observed

under repeated treatments, can be considered as an advantageous circumstance increasing the sensitivity of the experiment, and providing a control for differences between subjects. Each subject serves in this case as his own control, and the variability due to differences in the average responsiveness of the subjects can be eliminated from the experimental error. This is why experiments with "repeated measurements" can be handled as single-factor experiments in analysis of variance.

If the age span, during which the repeated measurements are carried out, represents a period of rapid developmental changes--as it was true during early conditioning in infants--the means of observations can change in terms of decimal orders. The mean rate of conditioning, for instance, can vary from several hundreds of trials in the youngest infants to several tens of trials in older infants, and corresponding changes can appear in standard deviations which show linear dependency on means. Then, adequate forms of transformation may become necessary for further statistical operations, such as a logarithmical transformation.

I think, that my contribution can be summarized as a description of a particular situation where a single investigation itself requires a longitudinal approach, and

where it is a conditio sine qua non. Problems connected with it can have more general importance, since in our case, they are problems of long term care for human infants at a special research unit which have not yet been thoroughly discussed in the literature. In this case, the longitudinal approach enabled us to perform a detailed analysis of the early development of some learning abilities, a detailed description of the course of conditioning in individual subjects, a separate analysis of age differences in those functions and of interindividual differences in a narrower sense. Hopefully, further observation of our subjects will ultimately elucidate the predictive value of these early differences.

LONGITUDINAL INQUIRY IN RETROSPECT

A. Harry Passow
Professor of Education
Columbia University, New York City

While we have used longitudinal methods in some of our work, I hardly consider myself "eminent" and I suspect there are those who would even question my being a "researcher." Well-designed, and well-executed long-term longitudinal studies are rather rare. Such studies as those of Terman, dealing with the gifted children; of the Bluecks, dealing with juvenile delinquency; of Nancy Bayley, on young children, and of the 1947 Scottish Council for Research in Education Studies (which are more basically developmental than educational in focus as I understand them) are some of the few examples we have.

More recently J. W. B. Douglas of England has reported findings from a study which began in 1946, dealing with the availability and effectiveness of pre-natal and maternity services, was followed up with a study of the health and illnesses of children to the point when they entered entrance school at age 5, and finally the development of the same children up to the point when they were eligible for leaving school at age 15.

Douglas' sample consisted of children born during the first week of March 1946. It included all of the children born to the wives of other kinds of manual workers and self-employed persons, beginning with 5,386 and ending some 15 or 16 years later with complete data on 3,418 and partial data on an additional 657 children. The sample is stratified for the entire population of England and Wales so that Douglas believes valid generalizations can be made. Beginning with maternity and prenatal care, the major focus in this third stage of the study has been on the educational growth of children and the consequences of certain educational practices. Therefore, I see this study, called The Home and the School, as being in essentially the same tradition as Terman, the Gluecks, and Bayley.

The importance of longitudinal research in education has to do with determining the stability of data and the depth of impact with respect to certain consequences and results of programs and procedures. For instance, I am convinced that one of the main weaknesses of much of the research that has been done on grouping stems from the fact that the time lapses have been inadequate for the test supplied. In other instances, I believe that some of the more significant results were missed because the time involved in research was

too short to have much of an impact, as such. In the latter category, for example, the psychological impact of desegregation may be such that the possible changes in academic achievement in other aspects of cognitive and non-cognitive growth are not revealed by the short-term testing which has been used on one study. The academic achievement of Negro youngsters who had been integrated was no better than it had been in the segregated situation. Study of self-concepts and ego development, tended to indicate that these were critical in this particular instance and until certain growth had taken place in these areas, very little change in academic achievement could be expected or found. Changes in self-concept and ego development take time and are not acquired by a three-week unit on improving one's self image.

Take another study, one on underachievement was done by the Talented Youth Project at Teachers College. Here, we had matched groups of underachievers with high potential but poor academic records. These boys were matched very carefully on a man-to-man basis and given two separate "treatments." At the end of the first semester, one might have concluded that the experimental program had had little effect, and in fact, the control group was doing better academically without special program modifications of any kind. However, at the

end of two semesters, the academic achievement of the two groups had reversed and now the experimental treatment was indeed having an impact on the underachievers. At the end of four semesters, both groups had "fallen apart", so to speak. There were a few "improvers" in both groups, but for most, there was relatively little improvement with respect to most of the boys involved. By the end of the third year (six semesters later), whatever improvements had taken place had pretty much evaporated and these youngsters did not achieve academically nor get the kind of school records that would have enabled them to use the potential which they apparently had and which they, indeed, had displayed through the first eight years of their schooling.

The point I am making is that if we take this three-year longitudinal study, depending upon the particular cut-off point, we would have drawn different conclusions and different inferences concerning the particular treatments and procedures applied. This, I think, is one of the major values of longitudinal research--it points out sharply the relative stability of findings, the artifacts that might occur over a shorter period of time, and lastly, the consequences of a developmental process, as such. Had we published our findings at any of the points--the end of one semester, or

two semesters, or four semesters, or six semesters--they would have been different and the value of a particular kind of grouping or of establishing particular kind of relationships or of stressing one kind of program over another would have come through quite differently at each of those points. We would have been more optimistic or more pessimistic about procedures at a particular time in this three-year study of underachievement. In this study we were intervening and using different techniques, procedures and materials, and would have drawn quite different inferences at each time interval.

Here is another example, a three-year study of mathematics for talented junior high school youngsters. In this study we were assessing the differential outcomes of mathematical achievement and attitudes toward mathematics of some six mathematics programs. We chose the three-year junior high school grades 7, 8, and 9. While I wouldn't want to defend this particular unit, I think there is some validity in our choice. We were comparing what started as six different programs or treatments. We started out with a total of 1,526 seventh graders in 51 classes, in 25 different school systems. We had eight classes following the SMSG (School Mathematics Study Group) as normally taught; six following SMSG accelerated; five beginning the University of

Illinois program two years early in seventh grade; and eight beginning the Illinois (UICSM) program one year early in the eighth grade and doing a so-called "modern math" program in seventh; a traditional mathematics program accelerated in ten classes; and 14 classes of traditional mathematics plus enrichment units. There is, of course, a difference in content to some extent, but there is also a difference in the sequencing or ordering of emphasis in each of these particular courses, so that if you assessed only at the end of six months or a year, it is quite possible that you would get a different kind of achievement patterning than if pupils were tested at the end of each of the years and at the end of the three-year period.

It is quite likely, of course, that if we were to continue to follow these youngsters through for three years more (grades 10-12) that the consequences of the particular junior high school program followed might have some impact as well.

A longitudinal approach is absolutely essential if one is to get insights into the overall changes in mathematics achievement, in ability to solve mathematics problems and in attitudes toward mathematics and not simply deal with the artifacts of a particular emphasis at a particular time.

I am assuming that the overall substantive differences in the programs are comparable to begin with, and this, it seems to us, is quite true from an analysis of the basic mathematics programs that we have been testing.

I also want to point out the care needed in setting up a longitudinal study. For example, in this Mathematics Study when we got the schools involved, explained the significance, the importance, and the nature of a commitment that would have to be made if they were to participate in the study, we made it clear that we did not want them to join unless they were ready and able to provide the kind of three-year commitment needed to keep the youngsters in the same class with the same group following the agreed upon mathematics program, allowing, of course, for normal attrition. We dealt with the superintendents of schools, the building principals, the heads of the math department, if there was such an individual, and with the teachers involved. All of these people were involved in the initial setting up of the study. We provided in-service training for the teachers in each of the programs involved.

We fed back data at the end of each year's testing program. From an initial population of 1,526 students at the beginning of seventh grade, we had lost only six at the end of seventh grade. We had an 11 percent loss at the

beginning of the eighth grade, and by the end of eighth grade were down to 1,357 students. It was at this point that we really had a major problem. Many of the schools that were involved in the so-called standard accelerated program, decided that the sequence they were following would mess up the taking of regents by their students. We could not convince them that it was perfectly possible to take the eleventh grade program for the tenth grade. The transition from the eighth grade to the ninth grades (especially when the youngsters were leaving a two-year junior high school and going to a four-year high school) was extremely traumatic for us, not for the kids. Although the problem of sequence and continuity for the entire three-year period had been made quite clear, we found that schools were unwilling to honor the commitment they had made. Some schools withdrew, leaving us high and dry at the end of two years; some insisted on changing the program. In effect, we had to drop them and modify the program. We were now dealing with a different set of dimensions than simply six different mathematics programs. We modified our design in such a way that we are comparing the problems of traditional or standard compared with contemporary math and accelerated with enriched programs.

The point I want to make is that by the end of the

ninth grade, instead of 51 classes, we were down to 38 classes. Instead of 1,526 students, we were down to 905 students; a 620 drop. Instead of having a minimum of five classes in any one of the programs being followed, we ended up with only one class in one of the programs and had to make other changes in the various cells we were filling. We still had a study, but it necessitated major adaptations and revisions in the appropriate statistical techniques which might be employed in the way we use the statistical procedures. There were major problems in the analysis of the data. I am trying to illustrate the idea that one ought have ~~no~~ no illusions about the complexities of doing longitudinal studies especially when such studies are to be done in school systems where one has no control over populations, or over programs, or over the idiosyncratic behavior of school administrators (if that's the proper term for it).

Another example is a two-year study we did of the affects of grouping on the academic achievement, the self-concepts, the attitudes toward more or less highly endowed peers and other consequences of broad and narrow range ability grouping in the fifth and sixth grades. Here, again, we involved some 80-plus classes with approximately 3,000 youngsters at the beginning of fifth grade. (I'm using this as an illustration of another kind of problem one runs into since these youngsters

were kept in intact classes and the New York City principals did basically honor their agreement with respect to holding the groups together). When it came to testing youngsters over the two-year period, we found that the academic achievement tests didn't "fall" the right way. By that, I mean, that if we used a particular form of the achievement tests, the floor was too high for the slower youngsters, and the ceiling was far too low for the more able youngsters. Two years later, having selected the test with the "lower floor" we discovered that the ceiling was far too low. The more able youngsters had "gone through the ceiling" at the beginning of the fifth grade, and then no growth movements at the end of the sixth grade. If we had interpreted our data at the end of the first year, we would have drawn different conclusions from those which we draw when we deal with increments of growth rather than with the absolute scores themselves.

I have been trying to point out some of the problems one encounters as he attempts to do longitudinal studies in the area of educational programs or curriculum studies. The virtues of longitudinal studies are no less important when dealing with consequences of educational programs over a period of time than they are when assessing the growth and development of individual children. But the problems, it

seems to me, are more complex in terms of population selection and sampling, of keeping track of such populations, of keeping the integrity of the design together, of finding appropriate measures, of applying what McClelland referred to as the "moving criterion" rather than "criterion specific." He argues (in his introductory chapter in Talent and Society), that criterion specific research may really not be intended to cope with problems of change, but rather for selecting talent potential for a specific objective. What he was questioning, is that in our immediate objective in selection on these bases, we may exclude a great many individuals who might perform better in terms of long-range and perhaps more important objectives. The same thing might apply in terms of longitudinal research generally; the application of multiple criteria over a period of time may have considerably more significance for us.

I have been reflecting on my own limited experience in longitudinal studies; the effects of ability grouping on fifth and sixth graders, the effects of various treatments on high school underachievers, the differences among various kinds of math programs with able junior high school youngsters. With respect to the latter study, we did a pilot project involving just four classes within a single school system over a three-year period. The coordinator was able to keep the

four classes intact, get complete data at the end of the seventh, eighth and ninth grades, and keep almost all of the sample available for testing at the end of the twelfth grade. Here, the findings at the end of the twelfth grade were essentially the same as those at the end of the ninth grade when we completed the major phase of our participation in this study. With regard to statistical or mathematical models or to the appropriateness to certain statistical techniques, I have available to me consultants. When I am confronted with statistical analyses problems, I go to people who are specialists and really know something and they help me out. Often I do not comprehend the techniques being applied but I rely on my consultants and my associates to carry the work forward. I know that there are researchers that think this is pretty bad and I wish that I had the time to become more skilled in procedures and techniques, but that is not possible now or in the foreseeable future. I see myself more as a "poser of educational problems", as a suggester of possible ways at getting at tests or analyses of alternative approaches; and as a coordinator of a research team which includes the kind of competencies needed for statistical design, for psychometrics for programming and computerization of data processing.

In short, I am not the person who should be teaching research and research design. Rather, I am a practitioner who is concerned that we do more research with children under somewhat messy and contaminated circumstances if we are going to get any kind of answers or leads that will be of value to us. This is my bias: it may be rationalization, but it is one I have to live with, nevertheless.

The barriers we face include the fact that we are dealing with fairly large and complicated samples in situations where we have to rely on the permanency or relative stability of the administration, of the staff, of the individuals who make commitments to us. We have to undertake our studies in such a way as to disrupt the on-going educational programs to the least extent. We have to rely on the goodwill of teachers, administrators, parents, and others so that they do not feel we are taking in inordinate amounts of teaching and learning time for testing. We have to find assessment techniques which will be valid and meaningful, and spread out over a period of time.

Some of the problems that I have been talking about are essentially inherent or caused by doing large scale research, dealing with large numbers of pupils and school personnel, in lifelike situations in an operating school and

are not, therefore, as much due to the fact of longitudinal research, as it is "school research."

I think that the basic problems are: keeping samplings intact, and getting assessment techniques that will really measure growth over the period desired and have the range for such measurement. I think that what we come out with frequently are really artifacts of the testing or assessment procedures we use and which lead us to generalizations which simply are not valid. Let me illustrate what I have in mind there. It is often said that disadvantaged children obtain lower scores on intelligence and achievement tests. Data often show that there is a progressive academic retardation and lowering of I. Q.'s scores as the youngsters proceed through the grades. Over and over again critics of schools are saying that this is caused by an educational deprivation, that is, the schools are contributing to the lack of achievement on the part of the youngsters.

Joseph Justman of New York City, studied the cumulative records of sixth grade pupils attending sixteen schools in disadvantaged areas in New York City. He found some 934 pupils who had entered the city schools in the kindergarten or first grade; had gone through the city school system without leaving or returning, moving in or out as many of these

youngsters do: had taken the Otis Alpha at the third grade level, the Otis Beta at the sixth grade level, and had taken the Metropolitan Reading Test at third and sixth grades. Now this, in a sense, is a longitudinal study using data retrospectively over a 6 or 7 year period. One other datum he had available was the total number of times that each of the pupils had been admitted to a different public school in the city. Of the total 934, 395 had attended only one school. The rest had moved as much as once, twice, three or four times. Some kids had moved six times during the first six grades. The interesting finding was that there was virtually no change in I. Q. from the third to sixth grade for the total group of 934 pupils. However, when this group was divided into sub-groups; a stable sub-group that had been enrolled in a single elementary school during their entire schooling and compare their performance with a group which had moved, then we find a very sharp and statistically significant difference. In the case of the stable sub-group, there is a mean rise of 1.6 I. Q. points. In the case of the mobile sub-groups, there is a mean drop of 1.3 I. Q. points. In both of these instances, the main difference, the third and sixth grade I. Q. was statistically significant. With respect to reading scores, these changes are even more startling as such, because Justman

concluded that the broad generalizations concerning changes with time (i.e., that the kids become progressively, academically retarded as they go through school) does not apply to the I. Q., certainly, and to the reading between third and sixth grade with a group of disadvantaged pupils who had not moved during that period of time. Mobility within the total educational situation may have a much more appreciable impact than other factors.

In Justman's study, there was a significant decrease in I. Q. observed only with the group of pupils who had attended four or more elementary schools. In the case of reading achievement, this group was also the only one which had significantly poor performance, as such. This represents a kind of longitudinal study that raises some highly significant questions about the academic aptitude and achievement of disadvantaged pupils. It relates attainment longitudinally to stability with respect to school placement rather than to the fact of "disadvantage" alone. Some of the conclusions that are drawn from it can be highly controversial (e. g., whether we should provide means for increasing the stability of the school experience as having greater significance as some of the other things we do. This is the kind of longitudinal research which is retrospective in nature.

COMMENTS ABOUT LONGITUDINAL RESEARCH

Robert F. Peck
Director, Research & Development
The University of Texas

My own experience with longitudinal research began at the University of Chicago in the eight-year study of adolescent character development known as the "Prairie City", "Midwest" or "Elm Town" study.

We went into it in the early 1940's with very few valid and workable guidelines as to the assessment of personalities, social role, or character structure. We had to improvise or even invent the measurement procedures. At that time projective techniques were actually disreputable. We found them among the most insight-giving kinds of data, and we had to provide ways of analyzing them which had not previously been worked out. The course of that particular study was educational, mostly by hindsight. We did not actually do an analysis of behavioral patterns in each year independently, then examine the course of development over the period of years included in the study. If I were doing it now, that is the way I would definitely choose to do it.

The next large scale study in which I participated was the Kansas City study of adult life. Chiefly for want of

funds, that had to be a cross-sectional study. This was particularly frustrating, since one of our major concerns was differential patterns of change during middle age and later life. By comparing age samples at age 45, 55, and 65, it was possible to identify some significant differences. But it was not possible to say whether these were developmental changes. Subsequently, a longitudinal study has been in progress with a new population in Kansas City, led by William E. Henry.

Since 1957, at the University of Texas, we have been conducting a series of longitudinal studies of personality changes during the college years. Some of these have been experiments to determine the differential effects of various kinds of educational or treatment procedures on student personality and behavior, over a period of one, two, or three years. The longest time we have data for, in large amounts, is three years. I'm particularly interested in a study of this period because of a hunch I have about the final firming-up of adult personality and behavior pattern at some rather specific point in time in the early twenties.

Longitudinal research follows individuals over a period of time. The time may be very short, measured in weeks, or it may extend over a period of years. Unlike all other kinds of designs, the focus on the individual makes longitudinal

research something qualitatively different from the purely statistical massing of data. It is possible to identify some uniquely different characteristics of each subject and trace what happens to these characteristics over the period of study, without contamination by data on other subjects. I believe this is an extremely important characteristic of longitudinal study and one which is insufficiently recognized and exploited.

Some theoretical assumptions underlying the longitudinal approach are, first, that development is sequential in nature and not symmetrical with regard to time. That is, as with the arrow of entropy in physics, the major stages are an inevitable product of the earlier stages, and could not simply be arbitrarily reversed without doing violence to the facts of life. A second assumption is that development proceeds through identifiable stages. While this may most safely be used as a heuristic device, it is of such long historical standing that it seems to me there is a good deal of validity to this way of viewing human development, as long as one does not get overly literal or over-simplify the great individual differences to be found in the age at which the stages are reached. A third assumption, which seems to me inevitable if one chooses to make longitudinal studies, is that there is a patterned consistency and stability of development and behavior over time, within the

individual life. This leads to a fourth assumption which I think is inherent in the choice of the method, and that is that individual behavior is predictable to a considerable degree. Or, to turn this around, one might say that the longitudinal approach is indispensable to the prediction of individual behavior. Indeed, in any situation where one tries to predict actions, whether by individual or grouped data, the very nature of the undertaking logically requires a longitudinal method.

In my own use of this approach all of these assumptions have come into play. I have been very much interested in both individual and grouped data which test the predictability of various kinds of behavior, ranging from moral behavior in adolescents to work output of adults. At Texas, we have been studying changes in mental health during the college years, both under naturally occurring conditions and when experimental interventions are used. In the latter studies, the emphasis has been almost wholly on the degree and kind of predictability possible. We have also been engaged in studies of the teaching behavior of some of these same subjects. Here, Dr. Francis Fuller and others have recently been developing a microscopic procedure for the analysis of classroom interactions, in which one important approach is the study of the sequences of teacher and child behaviors. If ever there was a longitudinal logic,

it certainly is essential to the very nature of such a study as this.

I cannot really say that any statistical model is uniquely appropriate to the longitudinal method. We have used just about every major form of mathematical model, including multiple regression equations to estimate predictive validity; factor analysis to identify stable dimensions to be followed over time; analysis of variance to determine the interacting effects, over several variables, at points through time; and hierarchial grouping analysis, a la Ward, to identify individual behavior or personality patterns which can then be differentially followed over a period of time.

As to the limiting effect of the longitudinal approach on statistical strategy, the only one I can think of is that the adding together of data on individual subjects is likely to do violence to the study of individual development. Consequently, if individual development is the question at issue, as I believe it inevitably must be for most questions of developmental psychology, then the usual methods of grouping data, searching for means and differences between means, are inherently inappropriate. But even this may be an overstatement, since it often is meaningful to compare two different groups of subjects as to their changes over time. This may do some violence to

the special properties of the individual developmental patterns, but where inter-group differences override intra-group differences, as they frequently do in some aspect of human behavior, then such classical statistical methods as analysis of variance, or other forms of measurement of group means and dispersions, are entirely appropriate.

As for the question of scientific classification, clearly the longitudinal approach is the method of choice--indeed, there is no other--if one wishes to classify different patterns of development through time. There are other contributions I could mention here, such as the identification of crucial periods in the life cycle, as Erikson's stages in psychological development, or Havighurst's development tasks for specified ages. Longitudinal study is not the only way to identify such stages, but it is by far the most pertinent.

As for the understanding of behavior, the longitudinal method has proved pragmatically indispensable. The historical reconstruction of experience in order to explain current behavior seems to be so meaningful and so necessary that almost no one in a psychotherapeutic experience forgoes it. In somewhat the same way, almost every theory which attempts to explain development invokes the sequential influence of early experience on

later developments, and even on the person's interpretation of later experiences, thus mediating their impact.

As for the construction and testing of theory, the observations I have made above about the uses, and the underlying assumptions of longitudinal approach, probably convey most of the ideas I could set down at this moment. In general, I can only reiterate that the longitudinal approach is logically indispensable to any theory of behavior change over time.

Moreover, the very construction of the word "predict" indicates the necessity for longitudinal measures in any predictive effort. As for control, clearly it is necessary to know what comes first, and what follows as a consequence, before one can have any assurance that intervention at a given point in time, will have a predictable, desired effect. Thus, a longitudinal strategy seems to me quite indispensable to the development of any high degree of predictive accuracy, or accurate control of development at a selected behavior period.

The disadvantages of the longitudinal approach are well-known. First of all, it inevitably leads to a high attrition rate in the experimental populations, particularly where these involve people who are free to move about in the natural course of events. For instance, in the Human Talent Study led

by Carson McGuire of The University of Texas, where some 1500 seventh graders were selected for study in four cities, by the 12th grade fewer than half this number could be located and retained as subjects in the longitudinal aspects of this research. This considerably increases the expense of such a study, for two reasons. First it is necessary to begin with much larger samples in order to have a minimal acceptable size left at the end of the study. Secondly, it is often much more expensive to locate and follow up subjects from the initial sample when they have moved to other locations, even within the same community.

Another well-known problem in longitudinal studies, is the fact that the ideas which shape the initial strategy may change during the course of a protracted research, to the point where the researchers become quite dissatisfied with their initial choice of methods. This is an inevitable price one has to pay, it seems to me. Indeed, some of our more fruitful advances in theory and technique have come precisely through such a process of exploring over a period of time. It may be a nuisance to a given research question, but overall it has greatly enhanced the power of research methods, in my experience.

Still another practical problem in the way of truly long term longitudinal studies, is the difficulty of maintaining the same research staff over a lengthy period of years. There

is probably some outside limit of 20 or 25 years which operates here in a practical way, even if it is not theoretically unavoidable. Research, after all, is done by dedicated individuals, not by anonymous, interchangeable units in a group. Consequently, good research is meaningfully pursued just as long as the original researchers are interested in it, or as long as they can personally enlist a few replacements to carry on for them.

So far, in the short history of behavioral science, there are few examples of survival after the first generation of researchers. A rare example is the Berkeley study of adolescent development.

As for the general promise of longitudinal research, it seems to me that I've probably already covered this. Suffice it to say that, with all of its practical difficulties, I've regarded it as the method of choice for anyone who proposes to establish the predictability of behavior, and for anyone who is interested in the formation of individual personality and its progressive reorganizations throughout the life cycle.

ON LONGITUDINAL INQUIRY

Hans Selye, M. D.
Institute de Medicine et de Chirurgie Experimentales
Universite De Montreal, Canada

It is true that the stress concept could not have been developed without a longitudinal approach since no other procedure could have detected the sequential development of the three basic stages: the stage of the alarm reaction, the stage of resistance, and the stage of exhaustion. In fact, it may well be said that almost all the individual parts of the mosaic constituting the general adaptation syndrome, or "G.A.S." were known long before my associates and I became interested in it and it was only the follow-up of the development of such changes in time that helped us to see the connections between what previously appeared to be unrelated observations.

However, perhaps the most characteristic feature of the "style" of research and postgraduate teaching in our Institute is the preference given to simple techniques as opposed to the currently fashionable complex technology of molecular biology. Consequently, we must sacrifice exploration of any one point in great detail, but the time gained permits us to accumulate a much greater number of observations for extensive correlations, either through

the longitudinal approach (to study the development of biologic phenomena in time), or for exploration of integrated functions in which many organs are involved at any one moment.

Let me point out that, like the vast majority of contemporary biologists, I am a great admirer of molecular biology; no one could fail to be impressed by the breathtaking new horizons made accessible through the application of modern complex physico-chemical methodology. Still, I feel that no matter how much we shall learn about minutia, there will always be a need for integrators who furnish an over-view of large fields. After thirty odd years of postgraduate teaching, it is my impression that a comparatively small proportion of Ph.D. candidates possesses the talents for correlative work, and in the life sciences there are comparatively few teachers of the old school left who are prepared to give them the necessary training.

In my book, From Dream to Discovery (McGraw-Hill, 1964), I have dealt with these problems at some length. There, my principal advice to budding scientists was: "Do not confuse the importance of your goal or the refinement of your tools with the significance of your work."

It would be a great pity if our craving to find the ultimate causes of things and our admiration for machines capable of demonstrating ever smaller individual building stones in matter would blind us to the advantages offered by the study--be it even on a more superficial level--of correlations such as I revealed by the longitudinal approach.

SOME IDEAS ABOUT LONGITUDINAL INQUIRY

Ruth Strang
Professor of Education
University of Arizona-Tuscon

At best, longitudinal research is complex and difficult to carry out. The simplest area of longitudinal research is the measurement of physical growth as represented by the Harvard Growth Study. There, the measures of height and weight were fairly standard and constant at different age levels. Less simple is the study of intellectual growth, for the measures of mental abilities which are far from comparable at the different age levels from birth to maturity.

Using the Harvard Growth data on intelligence, Robert Thorndike recently studied gains over several age periods. He pointed out the statistical problems and described the results as fragile. Still more complicated are developmental studies that attempt to show the growth in a relationship.

A second attempt was made by Harold Jones at the University of California to study the relation between the galvanic skin response and overt expression of emotion. By the longitudinal method, he was able to examine concomitant trends in internal emotional tension and out-going expressiveness from pre-school years to adult life.

The study of the development of the single relationship immediately suggests a much more complicated type of longitudinal study. There is no one-to-one relation between any two variables. Instead, there are many factors involved in the individuals interaction in a changing environment. Therefore, there is need to obtain data on many aspects of development physical, intellectual, emotional and social. This makes longitudinal studies alarmingly complex; yet, such complexity is necessary to achieve the most important goal of longitudinal research. The most important goal is not a description of physical and mental growth or growth in personality over a span of years. It is not even a study of a relation, such as the relation between personality and school achievement, as in the seven-year study at the University of Chicago. Without minimizing in the least the importance of these growth studies we must recognize the still greater significance of longitudinal studies that not only describe growth, but explain how that growth was achieved.

In 1958, Harold E. Jones recognized this extremely complex task of studying both intrinsic and external factors in the process of growth and aging. Gessel Institute investigations related the family environment to individual children's mental and personality development. The Guidance

Study at the Institute of Child Welfare, University of California, obtained data on a representative sample of several hundred Berkeley children during their first two years, and has followed them over three decades. From this monumental longitudinal study, as reported by Jean Walker McFarland, much has been learned about the factors which contribute positively to mature and effective personalities.

Studies such as these have prevented misconceptions arising from concealing differences in averages and assuming success for certain pre-school programs without studying the later consequences of the program.

There is still another aspect of longitudinal study in which I am now particularly interested. This is the study of sequential development of the reading process from pre-school pre-reading, pre-requisites to maturity in reading. It is a study of how individuals learn to read, how their reading competencies and interests grow, and what combination of conditions in the mental development, personality, family and neighborhood and school environment might explain different patterns of reading development. Past, present, and future are all involved. To study the learning process at any one point of time is difficult enough, but to study it longitudinally, presents so many research hazards that no prudent

investigator would even consider such an attempt.

The research hazards of longitudinal studies are well-known--the loss of subjects as the study proceeds, the difficulty of obtaining comparable measures on each age level, the development of valid measures, of many of the characteristics such as self concept, and value system which may be most influential in learning. The assumption is that these and other hazards may be avoided. Granted the significant data about the reading process could be obtained, the case study would seem to be the most appropriate method of treating the data. Comparable profiles and growth curves of related trends as well as statistical methods of treating the inter-related factors in the case study will have to be devised.

The emphasis in such a longitudinal study would not be on generalizations, for each individual is unique. It would not be on prediction for it is impossible to predict unless we know what the individual's future will be. Rather, the emphasis would be on such factors as describing the sequential development of the reading process, and on understanding the conditions that are favorable or detrimental to the development of the individual's reading potentiality.

A DIALOGUE ON LONGITUDINAL RESEARCH METHODOLOGY

J. Wayne Wrightstone

Assistant Superintendent, Board of Education, New York

Phillip Kraus

Professor of Education, Hunter College, New York City

Dr. Wrightstone: The New York City Board of Education has engaged in two major longitudinal research studies in recent years. In both studies no other research method was considered adequate for our purposes except the longitudinal research method. These studies were: (1) a study of elementary school children from kindergarten to high school graduation or high school leaving, a 13-year study to date, (2) a study of disadvantaged children with college potential in the demonstration guidance project from grades seven in the junior high school to college placement after graduation or school leaving before then, a seven-year study to date. In both studies we were interested in the relationships between the earlier and later changes in first, cognitive development; second, emotional and social development; and third, interests, aspirations and attitudes.

Dr. Phillip Kraus, now Professor of Education, at Hunter College, New York City, directed the study of elementary school children. He is continuing the longitudinal study. I'm going to ask him some questions about his long-range study or studies. Phil, what were the major objectives or emphases of

your longitudinal study of elementary school children?

Dr. Kraus: Well, you may recall that when we began we found that most studies as far as could be determined, were critically oriented or anthropometric, or psychological in nature in which they measured mental, motor, and physical development. As far as we could determine, there was hardly a study; in fact, there were no studies, which measured a large group of children over a period of time and observed growth and development in normal school situations. So we set up a study with no stated hypotheses but in an effort to answer questions such as: Are there discernable patterns in children's learning processes? How early does the level of achievement become fixed? How do current promotion practices of retention and acceleration affect children? How early is the promise of giftedness identified? How early are serious symptoms of maladjustment displayed? To what extent do traumatic experiences affect learning and adjustment? What is the significance of separation from parent involving anxiety and school phobia when starting school? And lastly, what is the effect of pupil mobility upon learning and adjustment? These, in general, were the objectives of the study.

Dr. Wrightstone: I can see that you have a very comprehensive list of objectives, and I realize now why this study

has taken so long. What types of pupil population did you study?

Dr. Kraus: We had hoped to study a group of pupils who are typical of the pupils of New York City, and were chosen at random. Although we did start with five schools, for a variety of administrative reasons and also reduction of staff, we ended up with a study of children in two schools. One was in a middle to low-middle income area near a middle income housing project. The other was in what is known as Harlem, which was a low socio-economic, but not the core of Harlem, not the lowest socio-economic level. In general, the population in the middle income area of Brooklyn was almost exclusively white; the population in a Harlem school was 91 percent Negro and about 5 percent Puerto Rican.

Dr. Wrightstone: Very good. What are some of the findings in cognitive development that were revealed by your longitudinal study?

Dr. Kraus: Suppose we speak first of intelligence and I. Q., because as you know, the measurement of intelligence by means of group tests has been eliminated by administrative fiat in New York City, but this followed our own study. Concerning intelligence, tests vary in the measurement of our children, and in the problem of trying to select tests for the

large off-city population the I. Q.'s that we had, were dependent upon the tests. Some of the tests used were very good for a Harlem population and decidedly not valid for our middle-income population. Other tests were better for the middle-income and not for the Harlem group. In general, the so-called deterioration of I. Q. that we have heard about, just didn't exist in our study; because the only deterioration of I. Q. that we found were in children whose I. Q. was above 110 in the first grade. This was equally true of our Harlem, low socio-economic population as well as of our white population. I might add that in no case did the I. Q. determine class placement, promotion, or retention, and that it was used only to help teachers in their planning for individual instruction. It's a very interesting finding, because we, in cross-section studies, have found that there seemed to be a deterioration in I. Q., so that this longitudinal study really provides us with some new information. Well, I think the public press and people, who know little about schools, exaggerate the importance of the I. Q. in the teacher's concepts of her children because teachers showed amazing independence from reliance on recorded I. Q.'s.

There were other aspects that we measured; for example, achievement, and this consisted mostly in the measurement of reading and mathematics achievement. As the study went on,

as children advanced through the grades, the relationship between reading and all other subjects became greater and greater. We found that in the early grades, certainly in reading readiness and reading achievement in the first and second grades, difficulties should be looked at with some interest and attempts to help made, but not with too much concern. By the third grade in this group, the pattern of reading tends to become rather definitely fixed. As we have gone on, we could have used third grade reading to predict almost everything that came after. Putting it another way, if you want to know how children will do in their I. Q. in sixth grade and ninth grade, look at third grade reading. If you want to know how they do in mathematics, and in all other subjects, again look at third grade reading. And at this point, I believe I haven't studied this fully, but I believe that there is a relationship between the College Board Scores of our college-bound children and third grade reading.

Dr. Wrightstone: Very interesting to have a finding such as that one. I feel pretty sure that only a longitudinal study would reveal facts such as these.

Dr. Kraus: We also did some studies of acceleration; and the Terman conclusions were fully verified. All the bright children who are accelerated at any time in their school, did

very well in their elementary school, junior high school, and senior high school, and were graduated a year ahead of the major group. I've seen a good many of them; they're getting along well in college.

Dr. Wrightstone: That again is a very interesting finding and particularly so in that it replicates the findings of Dr. Terman.

Dr. Kraus: We did a study on giftedness and talents and there the problem is a little bit mixed. We found out, for example, for the most part, special gifts and talents were discovered by teachers who possessed the gifts themselves, or at least to provide programs of such scope and variety that children have an opportunity to demonstrate their special talents. Of course, when programs of instruction were very narrow and limited, very few children were identified as being gifted. Nevertheless, there seems to be some consistency because children who are identified consistently as being talented or creative in music or art or dramatics or as being intellectually gifted, had been spotted for the most part before the end of the second grade.

Dr. Wrightstone: That's very significant to realize the early identification of these interests and gifts or talents.

Dr. Kraus: I should have mentioned along with acceleration our study of retention. There were, in fact, exactly 47 children who crossed our path who at one time or another had been retained or held back one year in a grade. Of the 47, as I recall, only two gained more than a year in their reading ability during the year of retention. Almost all of the others, gained less during the year in which they repeated the grade than they did in the many years in which they had been promoted. But I guess more serious is the fact that almost a quarter of the group were judged to have developed serious problems of emotional and social adjustment after having repeated the grade. And this, I think, is most serious.

Dr. Wrightstone: I agree with you, and this certainly confirms the opinion of many who feel that non-promotion is a poor method of attempting to improve the growth and development of the intellectual aspects of a child.

Dr. Kraus: The children, of course, are quite resentful; and the defensive mechanism that's set up would take too long to describe here. But I believe the damage was permanent with some of our children.

Dr. Wrightstone: I would agree with you. Now, do you have any findings that we haven't discussed with regard to the emotional and social development of these children?

Dr. Kraus: Yes. In this area we looked into problems of adjustment, problems of children who demonstrated considerable anxiety, or what we call separation from parent anxiety when they entered the kindergarten and school phobias. I might say, that in connection with general adjustment, we consider children to be maladjusted if at least three of their first seven teachers in their elementary school considered them to have problems of adjustment. If he was reported once or twice, we considered it not to be too important. It's interesting that of the children who were reported three or more times as having problems of adjustment in the elementary grades, almost all of them, that is all but three, had displayed their difficulties before the end of the third grade. And furthermore, they continued to have these same problems right through junior and senior high school. None of them grew out of it. Those who were identified after the third grades, sometimes did and sometimes didn't. But certainly the hard core of children with adjustment problems have been spotted sometime before the end of the third grade. This is very disappointing because we did not change the pattern.

Dr. Wrightstone: That is a particularly significant finding because in this area, I think we have much to learn particularly from the use of longitudinal method.

Dr. Kraus: I might mention that almost entirely unrelated, that the most sensitive predictor, if I may use that word, of maladjustment seems to have been the New York Reading Readiness Test, which was given early in the first grade, before most children had been identified as being seriously maladjusted. In fact, some of this group weren't identified until the second or third grade, and yet there was a most statistically significant difference in percentile ratings in favor of those who were eventually to become well adjusted as compared to those who were to have problems.

Dr. Wrightstone: This is a really interesting finding since we had not devised the Reading Readiness Test for this purpose.

Dr. Kraus: Well, there must be something in it that spotted this potential. There were other things we investigated; for example, we wanted to know the effect of severe traumatic experiences upon children and the experiences we isolated were really very severe. In general, no single traumatic experience seemed to alter the observable pattern of learning or of other behavior of our children. I would say that those who had traumatic experiences were able to cope with their problems and maintain their accustomed levels of performance at least in the school environment. They may have internalized it but

teachers could not identify any change in pattern.

We also did some examinations of both children who showed great difficulty in needing their mothers when first enrolled in school, and here the original group seemed to be slightly superior to the group of children who adjusted well. But as we went on through the years, neither intelligence nor subsequent achievement tests scores distinguished them from other children. I would say that the children who had difficulty leaving their parents, were really representative of the entire basic population. They included the subsequent successes, failures, the mature and well-adjusted, as well as those who are destined to be consistently identified as adjustment problems.

Dr. Wrightstone: There again, is a very interesting finding because this problem of separation at first grade or kindergarten level has not been explored too widely.

Dr. Kraus: I might add that in the high school we had two girls who developed really severe school phobias, to an extent that they had to be excluded from school and put on home instruction. These two girls were not among the group who showed this difficulty upon admission to kindergarten.

Dr. Wrightstone: Again, an interesting finding. Do you have any evidence of findings about the growth and changes

in aspirations or attitudes that were revealed by your longitudinal study?

Dr. Kraus: It still is a little too soon for this. We began with these children when they entered kindergarten. They are now seniors in high school or freshmen in college. It's interesting that some children that said in the first grade they hoped to be teachers, still want to be teachers in college; and some who said they wanted to go into medicine, still have the same aspirations. But some does not mean all, and there are many who have changed their aspirations, have changed their interests, and this holds a little bit more for our Harlem group, our Negro children than it does for our white children. I must say though, in determining parental aspirations both groups of parents seemed to have the same kinds of aspirations for their children. They were both interested in their children achieving as much as they could. We found almost equally, the statement, "I would like him to do whatever he can do well, and whatever makes him happy." These are almost the words that were used in both areas.

Dr. Wrightstone: That is really a rapid survey of all of the findings that I know you have in your manuscript. I presume that your study will be printed or published?

Dr. Kraus: Well, yes, the final revised manuscript, (I hope it's final) is now being examined by several publishers and I'll know shortly. I know it has been copyrighted and should be published, I hope, this year.

Dr. Wrightstone: I'm looking forward, with interest because I have read some of your manuscripts, and I know this study is going to find its place among the longitudinal studies that have been reported in educational literature.

Dr. Kraus: It'll be called Journey Through School.

Dr. Wrightstone: Well, Phil, I want to turn the tables now, and let you ask me some questions.

Dr. Kraus: Yes. I've been very much interested in the study of disadvantaged children with college potential and have never been close enough to it to know what it was about. Could you tell me something about the major objectives or emphases of this study on disadvantaged children with college potential?

Dr. Wrightstone: I shall be glad to give you a summary of our major objectives. First of all, Junior High School 43, Manhattan, and George Washington High School in New York City were the two schools through which we followed the children as they began their secondary education in Junior High School 43, Manhattan. Our major purpose was to demonstrate what could

be done to identify able students from socially and economically disadvantaged urban areas and to stimulate these able students to seek higher educational and vocational goals. Now this broad purpose can be broken down into some sub-problems or sub-objectives. There was, for example, the aim to determine whether the program that we were presenting in the curriculum and instruction in the schools resulted in the changes in scholastic aptitude or I. Q., to use a common term, of these project students. We were also interested to determine the effects on the program of academic achievement of these project students. We wanted to assess the aspirations and the attitudes of the project students, and also of their parents, as these were influenced by the program. We wanted to appraise the social adaptability of these students. This, in summary, is what I would define as major objectives of this study.

Dr. Kraus: This too, is a broad range of objectives. I wonder what kinds of pupils you worked with. What type of pupil population did you use?

Dr. Wrightstone: We used mainly economically and socially deprived or disadvantaged children. In Junior High School 43, for example, when we first began the study there, the school had approximately 48 percent of the children who were Negro, 40 percent of Puerto Rican parentage, and the

remaining 12 percent were white or others as we name them in our censuses in New York City.

Dr. Kraus: What are some of the findings, certainly in cognitive development, that your study revealed?

Dr. Wrightstone: Well, we have some interesting findings. Let us take first, the I. Q. or academic aptitude. As we stimulated these children to higher levels of study and application in their work, we found upon repeated testing with I. Q. tests that compared with the I. Q. of the eighth grade in the junior high school, by the time the children had reached the eleventh grade in the senior high school, we had differences in one grade group of about 13 I. Q. points increase, in another group about 10 points increase, and in a third group, again about 10 points increase. So that in this aspect of cognitive development we found a definite increase in the so-called I. Q. of these children. Now, with regard to the achievement of the pupils, when we first tested the pupils by a standard test in the seventh grade, we found that these pupils were approximately on the average, one year below the national norm. We re-tested these children, and by the time they had reached the tenth grade of the senior high school these children had increased their average development or growth in reading so that they were reading at the norm which

means that they increased in a period of about three years, by one additional year their actual achievement as a group in reading. We had a similar finding in terms of mathematics including arithmetic and algebra, where the children showed this growth. Now I do want to point out that the classes for these children were rather small, and in the case of mathematics, fifteen children to a teacher. English and Literature classes had a double period rather than the single period usually accorded the pupils from middle-class homes. So, we found that our program was having these effects on, not only the aptitude or I. Q., but also the academic achievement of the pupils.

Dr. Kraus: More intensive teaching and twice as much teaching created the results that you found?

Dr. Wrightstone: That is correct. It proved one of the assumptions that we started out with: that you can help these children from economically and socially disadvantaged homes to improve in terms of their academic aptitude and achievement.

Dr. Kraus: You must have done some work in the emotional and social development, too. Were there any significant findings revealed in these areas?

Dr. Wrightstone: Yes, we found some very interesting changes in these areas. We have the same guidance counselor stay with the same class, say the seventh grade. The seventh grade guidance counselor began with the class in the seventh, remained with the class in the eighth grade, ninth grade, tenth grade and the eleventh grade. And thus, we were able to follow the social and emotional development of the children. We found that the children were using the guidance counselor as a parent surrogate. They brought all kinds of family problems to the guidance counselor. The educational problems, if anything, were in the minority of the problems that the counselor had to discuss with them. But the counselor gave them such support that we observed that their emotional and social developments really showed gains during the period that they remained with us. And we have the children's own testimony in both solicited and unsolicited little writings that they submitted to us, in which they testify about the program and say that one of the major outcomes to them, personally, was the fact that it helped them emotionally, and helped them gain in their social development, their confidence in themselves, their ability to get along with others and similar attitudes that we usually associate with emotional and social development.

Dr. Kraus: This certainly points out the important role that a guidance counselor can play if he or she is programmed to function in a consistent way, in a longitudinal way moving on with these children. Our project assistant served the same role in the elementary school.

Dr. Wrightstone: Very good. This was one of the bonuses of our longitudinal study, because we did not realize when we started out that children from these disadvantaged homes would really grow into this close relationship with the guidance counselor, and I agree with you that this is an extremely important aspect. I'm just wondering, in fact, I believe, that children from more advantaged homes probably would also grow into the same relationship because we did find our children from our middle-income area doing it.

Dr. Kraus: Wayne, I wonder whether I might ask you the same question you asked me. Were you able to note any changes in interests and aspirations in attitudes among these children?

Dr. Wrightstone: Yes. This is an area where we did find great changes. The changes in interests in particular were very marked, because as we introduced them to various cultural activities in which they had never participated before, such as music, concerts, the theater and the like, we found

that children develop in interests that they had never had before; not only in these cultural things, but likewise in terms of educational and vocational interests. We found that they began to have aspirations in attitudes in which they, for the first time, said they wanted to go to college and thought they could qualify academically, in order to gain admission to these colleges. We found them developing interests in various types of vocations and professions as they undoubtedly were guided by teachers and guidance counselors so that I would say that through this longitudinal study, we really were able to observe very marked changes in interests, attitudes, and aspirations.

Dr. Kraus: I can't help remarking that it would be wonderful if this program were available to all the children in New York City, instead of just a few. How many were involved in your study?

Dr. Wrightstone: The total of the group which began in grades seven, eight and nine in Junior High School 43, Manhattan, but actually participated in the program in George Washington High School through to graduation was approximately 350 of the students. However, the project group itself numbered almost twice that number originally; namely, 700 when the program was initiated in Junior High School 43.

Phil, I am quite interested in the number of students or pupils that were involved in your study of elementary school children. I wonder if you could summarize that for me.

Dr. Kraus: Yes, you may remember that we began with all the children that entered the kindergarten in each of the two schools. There were approximately 300 at the time. We ended up at the end of the ninth grade with 80 children in each of the schools with whom we still have contact. Now of course, we had a staff through the sixth grade and only one member of the staff left through the ninth grade. And, since that time I've been doing it alone. I think, I know, there has been considerable attrition since, but I would say that I still have approximately 100 children now in their senior year in high school, or in college.

Dr. Wrightstone: Considering the fact that this is thirteen years, I think that's an excellent record. O. K. Phil, there is a second aspect of the longitudinal research method that we've been asked to react to. There is a list of questions that has been submitted to us and I thought we would discuss them rather informally. I'll read the first question and then we'll react to that. What are the salient characteristics of longitudinal research? What would you feel was one of the salient characteristics?

Dr. Kraus: Well, looking at the same individual year by year and noting change in progress gives us the kind of information that we can never get by getting a median or a mean of any single group in any single grade.

Dr. Wrightstone: Well, I would see another salient characteristic as an opportunity to study the interrelationships among the group components in the same individual. This was especially true as you were explaining some of your findings in your elementary school children as you study them in terms of cognitive and their social and other aspects of growth and development.

Dr. Kraus: Yes, and the impact of each of these phases on another in any single child can be observed only in a longitudinal study.

Dr. Wrightstone: Now, let's look at question number 2... What important theoretical assumptions are unique or indispensable to the longitudinal approach? What would you think of as one of these theoretical assumptions?

Dr. Kraus: Probably the most significance that we have seen in your discussion and certainly in my study, is the predictive value that a longitudinal study had because only, it seems to me in this way can we predict individual growth patterns or trace progress of any kind.

Dr. Wrightstone: Now, as I look at our studies, I see that this type of study is the only one that would permit us to compare the growth of brothers or sisters, or brother and brother in different classes. These longitudinal studies do permit what we call comparison of sibling growth at the same age, but then, of course, different times. So that I would feel that this is another one of the assumptions that we operate on.

Dr. Kraus: Yes, you have the interesting study really, in that you're at the same time doing a longitudinal study on changes in attitudes and beliefs and on feelings and hopes.

Dr. Wrightstone: Yes, then it does permit the developmental studies of an individual as a whole individual in a kind of social context and that, I think, is extremely important and cannot actually be done with any other research method except the longitudinal study. So it has real meaning.

Dr. Kraus: I would agree with you.

Dr. Wrightstone: Then there is the multi-discipline approach in which in most studies we are studying the growing individual including his physical, social, mental, and emotional developments. And to do this, we need that amounts to a multi-discipline approach.

Dr. Kraus: I would say that the lateral, the horizontal study, of any age group tend to wipe out, probably the most significant aspect of a child's growth, and wipe out some of the most important aspects of a child's growth, whereas, the longitudinal study is almost the only way of pin-pointing these things.

Dr. Wrightstone: Now, let's take a look at the third question... What role have these assumptions (that we've been talking about) played in your use of the longitudinal approach with elementary school children?

Dr. Kraus: Well, it's been the nature role; actually we worked on these assumptions and they've all been verified because the only conclusions we have been able to draw that were significant have certainly been those based on these assumptions.

Dr. Wrightstone: And I would say likewise in our study of the secondary school children, that we have pretty much made these basic assumptions, that we have really operated with these assumptions as an integral part, or the main part, of our approach through attempting to find out how these children did grow, develop, how their attitudes, interests, etc. changed, how their emotional and social development was affected. Now, what statistical or mathematical models would

you say are most helpful in employing the longitudinal approach?

Dr. Kraus: Really, the most meaningful methods are simple correlation and the development of growth curves based on these observations.

Dr. Wrightstone: Now, I would agree with you. I believe that these are the same statistical and mathematical models that we found most helpful in our studies of secondary school children. What is unique to the longitudinal approach which limits the use of certain common statistical techniques?

Dr. Kraus: I'm not so sure that anything limits the use. It's a matter of selecting the most useful and most valid statistical techniques. I don't see that there are real limitations in what was selected as being most fruitful.

Dr. Wrightstone: I would concur with you on your observation there, Phil. In question six, it says: In terms of the scientific hierarchy where can longitudinal study make the greatest contribution in: (a) Classify, describe, (b) Predict, control. Would you like to react to that, Phil?

Dr. Kraus: In terms of the last statement, the greatest contribution lies in the value of the longitudinal study in predicting and hopefully in controlling. It has not

been too successful, certainly in our group, we have not changed the prediction. But following that, it has offered material from which we can generalize and theorize. It's certainly helped us to understand and explain. The object of classifying and describing is a rather simple one, and this you can get from studies, other than longitudinal.

Dr. Wrightstone: Rather than repeat what you just said because it would apply with equal force in our study, I'm going to go on to the next question, and that reads: What major barriers have you faced in using the longitudinal approach, Phil?

Dr. Kraus: Well, they were all major. For one thing, the greatest one, is personnel. You recall that when we started we were doing two studies at the same time, and had only three full-time project assistants, in addition to myself. Then attrition took place, budgets were cut, and certainly from the sixth grade on, we had only one person other than myself, and since the Junior High School, we had no one but myself, and it's very difficult to follow children who are now in 28 different high schools. Secondly, money, because money begets data and puts it into usable form. But even more I envy the people who have projects whereby they can pay the receivers of questionnaires a dollar, or sometimes even five dollars

to return a questionnaire. Currently, my greatest difficulty is receiving back information from children who are no longer in school. The other aspect is time. You need a great deal of time. I could easily spend full time just being with the children and following them and speaking to them, or tracing some of them down. Another barrier, not really a barrier, we were really possessed with an urgency right through the study that every bit of data was terribly important. This is characteristic of all people who work on longitudinal studies for fear of omitting something. Although we did collect a great deal of data that we never used, I am very much pleased to know how much of it we did use. If I had to list them in order, I'd say money came first, personnel second, and time third.

Dr. Wrightstone: I would agree with you because we face some of the problems. Let's take the matter of personnel. Although there were several members of our staff who were able to be with us continuously for the seven-year period, we did have some turnover in the staff who were making the study so that it presents a problem of really orienting new personnel who come into the problem and into the types of data you're going to gather, so that personnel was one of the barriers

that we faced. Money, again, although we had some generous grants-in-aid in order to conduct certain aspects of this study as we approached the conclusion of the study, funds were running low and we had to cut some corners there and we did not gather as many data as we should have liked. As you pointed out, time is always a problem with those of us who are in research, including those of us who carry on a longitudinal research. Thus, we can not give full time to our longitudinal projects but have to face the practical problem of handling some other types of research and special studies also.

Dr. Kraus: I might add that the Board of Education supported this study for six years and a little for the next three years. But it would have stopped dead in its tracks had it not been for Hunter College, and a foundation which came through with grants. But this could stop any moment.

Dr. Wrightstone: I think you were very fortunate to obtain this support when the Board of Education was not able to supply you with funds to continue your follow-up study. What would you say is the general promise of longitudinal research as we look at it, Phil?

Dr. Kraus: I believe that it can contribute and give a quality of data which is completely lost in other studies

which become qualitative. Certainly with many of our children, the qualitative aspects of the study have been so significant for their growth and so significant for us in interpreting what has happened to them, and almost in predicting what can happen. I think the general promise of longitudinal research is the one that in the study of children probably has greatest validity.

Dr. Wrightstone: I would agree with you, Phil. I jotted down a few of the studies as I could recall them, which had really been a bulwark for us in research. Going back to B. T. Baldwin's study of physical characteristics of children; here the longitudinal method has given us something that no other type of research method or study could have provided. The same is true of Dr. L. M. Terman's studies of gifted children where he has followed them in some instances for periods of several decades: Arnold Gessell's famous work on Infant Growth and Development; W. F. Dearborn's Harvard Growth Study which is often cited; and one study with which I was associated in its early years was the study directed by Dr. Ralph Tyler, popularly known as the Eight-Year Study. These it seems to me, are indicative of the general promise of the longitudinal research method in that in the future equally significant studies will be made using this method.

Dr. Kraus: I might add Havinghurst's study which was published in 1962, called "Growing up in River City," and what attracted me particularly is that it begins at the point where my current study, or that is the reported part of the current study, and that is at the end of the high school level.

Dr. Wrightstone: Well, this has been very kind of you Phil, to sit down here and talk with me because I always have a fear of talking alone to myself; and it's been much better that we could carry on a little dialogue here.

A BIBLIOGRAPHY OF
MEASUREMENT OF GROWTH; FUNCTION FLUCTUATION

- Anderson, T. W. "The Use of Factor Analysis in the Statistical Analysis of Multiple Time Series." Psychometrika, 1963, 28, Pp. 1-25.
- Bock, R. D. "Multivariate Analysis of Variance of Repeated Measurements." In C. W. Harris (Ed.), Problems in Measuring Change. Madison: University of Wisconsin Press, 1963, Pp. 85-103.
- Hoffman, E. L and L. R. Tucker. "Three-way Factor Analysis of a Multitrait-multimethod Matrix." Urbana: University of Illinois, Department of Psychology, Technical Report, 1964. (Mimeographed)
- Holtzman, W. H. "Methodological Issues in P Technique." Psychology Bulletin, 1962, 59, 248-256.
- _____. "Statistical Models for the Study of Change in the Single Case." In C. W. Harris (Ed.), Problems in Measuring Change. Madison: University of Wisconsin Press, 1963. Pp. 199-211.
- Levin, J. "Three-mode Factor Analysis." Urbana: University of Illinois, Department of Psychology, Technical Report, 1963. (Mimeographed)
- Potthoff, R. F. and S. N. Roy. "A Generalized Multivariate Analysis of Variance Model Useful Especially for Growth Curve Problems." Biometrika, 1964, 51, Pp. 313-326.
- Tucker, L. R. "Experiments in Multi-mode Factor Analysis." Proceedings of the 1962 Invitational Conference on Testing Problems. Princeton, New Jersey: Educational Testing Service, 1962.
- _____. "Implications of Factor Analysis of Three-way Matrices for Measurement of Change." In C. W. Harris (Ed.), Problems in Measuring Change. Madison: University of Wisconsin Press, 1963, Pp. 122-137.

Tucker, L. R. "The Extension of Factor Analysis to Three-dimensional Matrices." In N. Frederiksen & H. Gulliksen (Eds.), Contributions to Mathematical Psychology. New York: Holt, Rinehart & Winston, 1964, Pp. 110-127.

Wiggins, R. A. "On Factoring the Correlation of Discrete Multivariable Stochastic Processes." Massachusetts Institute of Technology, Cambridge 39, Massachusetts, Scientific Report No. 9, 1965, 196 pp.

Submitted by:

Frederic M. Lord, Chairman
Psychometric Research Group
Educational Testing Service
Princeton, New Jersey